A Study of Mammalian MicroRNA-Mediated Repression of Gene Expression by Ribosome Profiling

by

Huili Guo

B.A., Natural Sciences (2005)
University of Cambridge

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2011

Signature of Author.....
Huili Guo
Department of Biology
May 20, 2011

Certified by.....
David P. Bartel
Professor of Biology
Thesis Supervisor

Accepted by.....
Robert T. Sauer
Luria Professor of Biology
Chair, Biology Graduate Committee

A Study of Mammalian MicroRNA-Mediated Repression
of Gene Expression by Ribosome Profiling

By

Huili Guo

Submitted to the Department of Biology on May 20, 2011
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

ABSTRACT

All cells in a multicellular organism carry the same genes, yet these same genes direct the differentiation of many different cell types. This is facilitated by differential gene expression, the control of which can be exerted at the transcriptional, as well as post-transcriptional, level. MicroRNAs (miRNAs) are ~22-nucleotide small RNAs that mediate post-transcriptional regulation of gene expression by base pairing to their target mRNAs to direct repression. In animals, this repression is usually mediated through translational repression and/or mRNA destabilization.

In studies that investigate miRNA-mediated repression with reporter constructs or individual endogenous genes, translational repression and mRNA destabilization have been observed to contribute variably to the overall level of repression. This led to the question of whether the same was true for endogenous targets at a genome-wide level. While changes in mRNA levels can be easily captured by microarray measurements, it is harder to measure translational repression on a genome-wide scale. To address this gap, we used ribosome profiling to measure effects on protein production and compared these to simultaneously measured effects on mRNA levels. The latter were also quantified by a deep-sequencing approach (mRNA-Seq). This enabled us to obtain a snapshot of changes in translational efficiency at the genome-wide level. For both ectopic and endogenous miRNA regulatory interactions, we observed that lowered mRNA levels account for most (≥84%) of the decreased protein production. These results show that changes in mRNA levels closely reflect the impact of miRNAs on gene expression and indicate that destabilization of target mRNAs is the predominant reason for reduced protein output. The slight reduction in translational efficiency is likely mediated by an inhibition of translation initiation.

For studying miRNA repression in an endogenous system, we had initially used *in vitro* differentiated neutrophils from *mir-223* knockout mice and compared these to cells from wild-type mice. Because neutrophils have a shorter lifespan than most differentiated cell types, we selected another endogenous system, antigen-stimulated B cells from *mir-155* knockout mice, and similarly compared these to cells from wild-type mice. In addition to mRNA-Seq and ribosome profiling, we made parallel proteomics measurements. Our results show that miR-155 in antigen-stimulated B cells primarily mediates mRNA-level changes, though the contribution from changes in translational efficiency was larger than previously observed. In addition, we observed widespread translation of upstream open reading frames initiated from canonical and non-canonical start codons. These upstream open reading frames are also translationally repressed by miR-155.

Thesis Advisor: David P. Bartel
Title: Professor

Acknowledgements

Six years ago, I first came across microRNAs in my last year of undergraduate study, and became deeply intrigued. I never would have imagined that one day I would be contributing something back to the field. For this, I have many people to thank.

I am immensely grateful to my advisor, Dave Bartel. It is a great pleasure to be able to work with such a brilliant mind. He has taught me how to think critically, logically, and be meticulous to a fault. To say that I have learnt a lot from him would be a massive understatement. Most of all, I thank him for his patience when I was picking up computational biology, and his belief in me throughout the years.

I thank my committee members, Phil Sharp and Tom RajBhandary. Thank you for your time and expertise, your patience with me. I could not have asked for a better committee.

I thank everyone in the Bartel lab. They have made lab life (and out) extremely edifying and enjoyable. I have to especially thank Jinkuk, Sue-Jean, David G., Anna, Jin-Wu and Vincent, who have spent many hours listening to my joys and frustrations, be they willing or unwilling when I accosted them in lab. Special thanks to my two baymates, Mike Lam and David W.: my bay has become infinitely more happening because of you. Thanks also to Rosaria, my fellow gel-room-keeper: I wouldn't have known how to manage without you.

I thank my collaborators Nick Ingolia and Jonathan Weissman, especially Nick, for his selfless sharing and invaluable advice. I have learnt and benefited a lot from his brilliance. I also thank Chanseok, Daehyun and Judit for their help with the project I am working on now.

I am grateful to Chris Petersen. The topic of my project, his graduate work experience and convenient postdoc location in the Reddien lab meant that my asking and his giving were always going to be on a collision course. And he has been more than generous with his time.

I am thankful for having known Calvin and Graham. They are the two people whose brilliant minds and intellect I have always admired and they taught me what being selfless really means. Their encouragement, advice, and their friendship is something I will always treasure.

I have made a very special friend in my time here — Kathy. From our times together in class, in lab, at ice cream sessions, and now emailing sessions, she has always been a great source of support. Her inventive mind is a sounding board for my half-baked ideas. Her dedication to science both inspires and humbles me.

I thank my funding source, the Agency for Science, Technology and Research (A*STAR). If not for them, I would not have been at Cambridge for undergrad, and would never have had the chance to be at MIT. I am truly grateful. I also thank my friends in the Singaporean community at MIT, for the food, singing, fun and laughter.

I thank my scientific mentors and friends, Ben McCaw, Fiona Wardle, Richard Farndale, Sumana Chandramouli and Graeme Guy. They saw something in me that I did not see years ago, and are directly responsible for many of my scientific idiosyncrasies, right down to the way I keep my lab notebooks.

I thank my parents and my brother, for being so wonderfully supportive in all the years that I have been away. The Saturday morning calls home work wonders when the going gets tough, and my brother's innocuous comments never fail to remind me that there is an entire world outside of my grad school bubble.

Lastly, I thank my better half, Vincent. From our Cambridge days, whether he was here, or far away, he has been my greatest supporter and my strongest motivator. I thought I was a workaholic until I married one. But I am a better person for having him in my life. Thank you, and love always.

Table of contents

Abstract 3

Acknowledgements 5

Table of contents 7

Chapter 1. Introduction 9
- History of RNA interference 9
- History of the discovery of microRNAs 12
- Other classes of endogenous small RNAs 14
- MicroRNAs: Biogenesis 15
- MicroRNAs: Target prediction 17
- MicroRNAs: Function 19
- mRNA stability 21
- Regulation of mRNA stability 22
- Eukaryotic translation 22
- Study of translation 24
- Regulation of translation 25
- Interplay between translation and mRNA stability 29
- Interaction between the 5′ and 3′ ends of an mRNA 30
- Molecular consequences of microRNA-mediated repression 32
- Figure legends 38
- References 41
- Figures 52

Chapter 2. Mammalian microRNAs predominantly act to decrease target mRNA levels 55

Chapter 3. A genome-wide study of miR-155-mediated repression in activated B cells 95

Chapter 4. Future directions 133
- Implications for microRNA-mediated repression 133
- Repression of upstream open reading frames 134
- Repression mediated by sites in the open reading frame 135
- Primary event of miRNA targeting: first translational repression, or first decay? 136
- Stress and the balance between translational repression and mRNA destabilization 137
- Stress and its potential effect on mRNA stability 139
- Stress and its potential effect on translational efficiency 141
- More translational repression, or more decay — does it matter? 143
- Concluding remarks 145
- Acknowledgements 146
- References 146

Appendix. The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs 149

Curriculum Vitae 155

Chapter 1

Introduction

In Jacob and Monod's 1961 classic review article, two models of gene expression regulation were proposed: transcriptional control (Model I) and post-transcriptional control (Model II) (Jacob and Monod, 1961). In both models, an unstable intermediate contains information that has been transcribed from stable DNA genes; this information can then be translated into a unit of function, the protein, by ribosomes and other components of the protein synthesis machinery. A repressor can either repress expression by acting at the site of the DNA genes (Model I) or by acting on the messenger itself (Model II). Shortly after, the "messenger hypothesis" was confirmed in bacteria (Brenner et al., 1961; Gros et al., 1961). Over the years, however, post-transcriptional control has played more of a second-fiddle role to transcriptional control[1]. Post-transcriptional control, which includes the control of translation, continued to be studied, but mainly in systems that necessitate this occurrence; for example, in red blood cells that have enucleated, and in fertilized eggs of invertebrates, which were known to be transcriptionally silent for several hours right after fertilization (Gross et al., 1964).

In the last few decades, however, post-transcriptional control has gradually come to the fore, even in nucleated cells that are carrying out active transcription. This is especially so with the discovery of RNA interference and its link with endogenous pathways.

History of RNA interference

In 1990, in an attempt to enhance the purple coloration of petunias, transgenes encoding a pigment-producing enzyme, chalcone synthase, were introduced into the plants. Unexpectedly, the overexpressed

[1] Bacterial systems were much better studied at the time; thus perhaps because bacterial mRNAs are much less stable than eukaryotic mRNAs, and there is no compartmentalization between transcription and translation in bacteria, the idea of post-transcriptional control did not find much traction, when compared to transcriptional control.

transgene resulted in less pigmented, not darker, flowers (Napoli et al., 1990; van der Krol et al., 1990). Not only were the transgenes silenced, endogenous genes containing homology to the transgenes were also silenced. Because the expression of endogenous genes could be altered by exogenous elements, the term "cosuppression" was used to describe this phenomenon. The silencing mechanism employed in cosuppression in plants can be transcriptional[2] or post-transcriptional; the latter, known as post-transcriptional gene silencing (PTGS), was also well-studied in *Neurospora crassa*, where it was known as "quelling" (Cogoni et al., 1996; Cogoni and Macino, 2000; Romano and Macino, 1992).

Meanwhile, in the *Caenorhabditis elegans* community, efforts were made to probe gene function by injecting antisense RNA targeting the gene of interest[3]. While antisense RNA injection led to the expected phenotype, the control experiment — injection of sense RNA — resulted in the same phenotype, with the same level of penetrance (Guo and Kemphues, 1995). It was not till three years later, when Fire and Mello reported that injection of double-stranded RNA (dsRNA) causes a much stronger phenotype was this curiosity explained. It turns out that the production methods used had resulted in dsRNA-contaminated preparations of antisense RNA and sense RNA. When this contaminating dsRNA was purified away, most of the silencing activity was lost (Fire et al., 1998). That dsRNA was such a potent and specific inducer of silencing made this an extremely exciting observation, and this new phenomenon[4] was dubbed "RNA interference (RNAi)".

---

[2] Transcriptional gene silencing (TGS) is a related phenomenon, in which gene expression is blocked at the DNA level, usually by DNA methylation, and was first observed in plants (Wassenegger et al., 1994).

[3] The prevailing view at the time was that an antisense RNA complementary to its intended target could bind to the target mRNA and inhibit translation of the target mRNA.

[4] In hindsight, a similar phenomenon had already been observed in plants infected with viruses. In one study, when plants were infected with potato spindle tuber viriods (PSTV) containing fragments of an endogenous gene, the endogenous gene itself was silenced by TGS (Wassenegger et al., 1994). In another study, instead of getting high levels of transgene expression through amplification of viral RNA by the viral RNA-dependent RNA polymerase (RdRP), plants transformed with potato virus X complementary DNA (cDNA) only had moderate transgene expression, coupled with viral resistance (Angell and Baulcombe, 1997). Genes with sequence homology to the viral RNA were also silenced, much like PTGS. In both studies, replication-defective viral sequences failed to generate the same response. As the viral replication cycle generates dsRNA, this suggested that dsRNA was involved in the silencing mechanism. Even earlier, in 1986, Beachy

By 1999, Hamilton and Baulcombe had demonstrated the first association between small RNAs and PTGS in plants (Hamilton and Baulcombe, 1999). Soon after, dsRNA-derived small RNAs were linked to gene-silencing phenomena in embryonic cell extracts (Hammond et al., 2000; Zamore et al., 2000). A combination of genetic and biochemical approaches then identified the main players involved in turning dsRNA into small RNAs capable of silencing target RNAs.

Double-stranded RNA is processed into short ~22-nucleotide (nt) RNAs by an RNase III enzyme, Dicer, in a phased manner measuring from one end of the dsRNA (Bernstein et al., 2001; Elbashir et al., 2001a; Ketting et al., 2001; Zamore et al., 2000). This generates small RNA duplexes, known as small interfering RNAs (siRNAs), with ends that are characteristic of RNase III cleavage — namely, 5′ phosphates and 2-nt 3′-hydroxyl overhangs (Elbashir et al., 2001a). These siRNAs then get incorporated as single-stranded RNAs into a ribonucleoprotein complex known as the RNA-induced silencing complex (RISC)(Hammond et al., 2000; Martinez et al., 2002; Nykanen et al., 2001). The incorporated siRNA serves as a guide, providing the specificity to target mRNAs that are perfectly complementary to the guide strand (Bernstein et al., 2001; Elbashir et al., 2001a; Zamore et al., 2000). Guide strand selection is dictated by the thermodynamic stability of the ends of the siRNA duplex. The strand that enters RISC is almost always the one whose 5′ end is less tightly paired (Khvorova et al., 2003; Schwarz et al., 2003)[5]. Cleavage of the target is then mediated by the endonuclease activity of Argonaute (Ago), the core component of RISC (Liu et al., 2004; Song et al., 2004). The target mRNA is cleaved in a characteristic manner — always at the position between the nucleotides pairing to bases 10 and 11 of the siRNA (Elbashir et al., 2001a; Elbashir et al., 2001b).

---

and colleagues had introduced a transgene encoding the tobacco mosaic virus coat protein into tobacco plants and observed increased resistance to viral infection (Abel et al., 1986). The prevailing idea then was that coat protein overexpression interferes with virion disassembly in initially infected cells, thus blocking the spread of infection; this phenomenon can now be rationalized by RNA-mediated silencing mechanisms.

[5] The high reliability of this rule serves as a very convenient rule-of-thumb for biologists designing siRNAs for their targeting experiments.

The parallels in plants (PTGS), fungi (quelling), worms, flies, and mammalian cell extracts (RNAi) suggested that the components that make up this new, almost-too-good-to-be-true gene-silencing tool was evolutionarily conserved. Indeed, Dicer and Argonaute proteins are both phylogenetically widespread, with their respective domains deeply conserved across the evolutionary tree (Cerutti and Casas-Mollano, 2006). By the late 1990s, gene-silencing phenomena had been linked to the silencing of viruses and transposons in plants and nematodes (Anandalakshmi et al., 1998; Beclin et al., 1998; Brigneti et al., 1998; Kasschau and Carrington, 1998; Ketting et al., 1999; Tabara et al., 1999), but a biological role had not yet been demonstrated in mammals; surely, the protein components in this pathway had not been selectively maintained in mammals just for the convenience of future biologists?

History of the discovery of microRNAs

Unbeknownst to most researchers at that time, part of the answer to that question had already been found, in *C. elegans*. With its well-characterized cell lineage (Sulston and Horvitz, 1977), *C. elegans* had been used extensively in the study of developmental timing. Worms incapable of exhibiting normal developmental timing are known as heterochronic mutants. One pair of such mutants was *lin-4* and *lin-14*. *lin-4* loss-of-function mutants exhibit inappropriate reiterations of early cell fates at late developmental stages (Chalfie et al., 1981). *lin-14* null mutants, on the other hand, exhibit precociously early cell fates and are epistatic to *lin-4* nulls, suggesting that *lin-4* is a negative regulator of *lin-14* (Ambros, 1989). Another mutant allele of *lin-14* (*lin-14* gain-of function, or *lin-14(gf)*) led to the realization that the *lin-14* 3′-untranslated region (3′UTR) was necessary for this negative regulation (Wightman et al., 1991). In 1993, *lin-4* was found not to encode a protein, but a 22-nt RNA with antisense complementarity to multiple sites in the 3′UTR of *lin-14*, which, together with reporter fusion experiments, suggests that the binding of *lin-4* RNA to the *lin-14* 3′UTR led to the downregulation of *lin-14* expression (Lee et al., 1993; Wightman et al., 1993). In particular, when *lin-4* was present, LIN-14

protein levels were downregulated, but *lin-14* mRNA levels were not found to be substantially changed, indicating that *lin-4* binding somehow represses *lin-14* translation[6].

This oddity was found not to be an isolated example, when another small RNA, *let-7*, was discovered, and also targeted genes involved in developmental timing (Reinhart, 2000). Moreover, *let-7* was phylogenetically conserved across bilaterally symmetric animals (Pasquinelli et al., 2000). These small RNAs that regulate developmental timing were christened short, temporal RNAs (stRNAs). It was soon discovered that stRNAs were not oddities in themselves either, at least not in size. By 2001, more small RNAs were identified in worms, flies and mammals (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001), forming a new abundant class of RNAs called microRNAs (miRNAs), for which *lin-4* is the founding member.

By this point, the discovery of abundant miRNAs was taking place contemporarily with the rapid developments in the RNAi field, and the parallels between the two types of small RNAs and their processing pathways were hard to miss. miRNA processing also requires Dicer (Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001), and when miRNAs were found to associate with a ribonucleoprotein complex, the miRNA ribonucleoparticle (miRNP) which has very similar components to the RISC involved in RNAi (Mourelatos et al., 2002), the immediate question prompted was whether these small RNAs of different origins could act in the same way. While miRNAs were not originally thought to cleave their target mRNAs, as siRNAs do, it was soon discovered that a miRNA can also specify cleavage of a reporter RNA as long as the target has extensive complementarity to the miRNA (Hutvagner and Zamore, 2002). Conversely, siRNAs that were originally designed to cleave their targets

[6] Note that this is remarkably similar to Model II, as posited by Jacob and Monod, and was exactly the premise behind the antisense RNA injections that led to the discovery of RNAi.

could also repress gene expression in a miRNA-like manner, when complementarity between the siRNA and its intended target was insufficient to direct cleavage (Doench et al., 2003; Zeng et al., 2003)[7].

Other classes of endogenous small RNAs

In addition to miRNAs, the advent of high-throughput sequencing has led to the discovery of other classes of endogenous small RNAs at a breakneck pace. One intriguing class of endogenous small RNAs delivered by the deep-sequencing juggernaut are the piwi-interacting RNAs (piRNAs)[8]. piRNAs are germline-specific small RNAs that are ~24–31 nt long. First described in *Drosophila* as repeat-associated small interfering RNAs (rasiRNAs)(Aravin et al., 2003), piRNAs have since been found in mammals (Aravin et al., 2006; Girard et al., 2006; Lau et al., 2006), zebrafish (Houwing et al., 2007), and even in *Nematostella* and *Amphimedon* — animal phyla that diverged before the emergence of bilaterian animals (Grimson et al., 2008). At least a subset is derived from repeat elements and hypothesized to be generated by a fascinating ping-pong mechanism (Brennecke et al., 2007; Gunawardane et al., 2007). This process contributes to the silencing of retrotransposons, constituting a genome defense mechanism that is especially important in germline development (Siomi et al., 2011).

Yet another class of small RNAs that has recently been identified through deep-sequencing approaches is a group of endogenous siRNAs (endo-siRNAs). These were first discovered in plants (Llave et al., 2002; Reinhart et al., 2002), fungi (Reinhart and Bartel, 2002) and nematodes (Ambros et al., 2003; Ruby et al., 2006), and for a long time, endo-siRNAs were thought to be specific to such

[7] Note that this only applies to Ago proteins that are capable of mediating cleavage. For example, in humans, there are four Ago proteins (Ago1–Ago4), all of which can carry out target repression, but only one can specify cleavage. The PIWI domain of Ago proteins has an RNase H-like fold, and like RNase H-like enzymes which cleave RNA using a DNA template, catalysis requires a catalytic triad, usually made up of Asp-Asp-His (or DDH). Although the DDH catalytic triad is necessary for catalysis, it is not sufficient — both Ago2 and Ago3 harbor the catalytic triad but only Ago2 is able to catalyze cleavage (Peters and Meister, 2007).

[8] The Argonaute family of proteins can be divided into two clades: the Ago clade, which predominantly interacts with miRNAs and siRNAs, and the Piwi clade, which interacts with the piRNAs.

organisms, which possess RNA-dependent RNA polymerases (RdRPs)[9]. As most endo-siRNAs are produced via the action of RdRPs, flies and mammals were not expected by some to have endo-siRNAs (Ghildiyal and Zamore, 2009). However, recent studies have reported the presence of just such small RNAs in flies and mammals, implying that endo-siRNAs can also be generated without RdRPs (Babiarz et al., 2008; Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008a; Okamura et al., 2008b; Tam et al., 2008; Watanabe et al., 2008). These endo-siRNAs are derived from genomic sources of dsRNA triggers such as loci with a high degree of secondary structure, and convergently transcribed loci (Ghildiyal and Zamore, 2009). The biological function of endo-siRNAs is still unclear, though it is intriguing that in oocytes, one of the sources in which endo-siRNAs have been sequenced, long dsRNA does not activate the interferon response (Stein et al., 2005); oocytes might thus represent a permissive environment for endo-siRNAs to regulate endogenous transcripts.

MicroRNAs: Biogenesis

Once loaded into miRNP/RISC, miRNAs and siRNAs are biochemically indistinguishable but they do differ in terms of their origin. Mammalian miRNAs can be encoded as independent transcription units (monocistronic), as a cluster with other miRNAs in a single transcription unit (polycistronic), or within an intron/exon of other protein coding genes (Kim et al., 2009). They are typically transcribed by RNA polymerase II (Pol II). Once transcribed, the primary transcript (pri-miRNA), which has local stem-loop structures, is processed to liberate the pre-miRNA in the form of a small hairpin of ~60–70 nt (Figure 1A). This processing takes place in the nucleus, and is performed by another RNase III enzyme, Drosha, together with its cofactor DGCR8 (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Landthaler et

[9] This is also the reason RNAi in plants and worms can be amplified (Ghildiyal and Zamore, 2009). In plants, a target mRNA, upon cleavage, serves as a template for RdRP to synthesize the complementary strand, giving rise to new sources of dsRNA that can be processed into secondary siRNAs; in worms, RDE-1, the Ago protein loaded with primary siRNAs, guides RdRP to the target mRNA, which serves as a template for the synthesis of secondary siRNAs, without requirement for Dicer.

al., 2004; Lee et al., 2003). The Drosha:DGCR8 complex cleaves the pri-miRNA ~11 base pairs from the base of the hairpin stem (Han et al., 2006) and sets the register of one end of the eventual small RNA duplex. The pre-miRNA is then exported out of the nucleus by Exportin-5 (Lund et al., 2004). In the cytoplasm, the PAZ domain of Dicer recognizes the 2-nt 3′ overhang of the Drosha-generated end, just as it would one end of a long dsRNA. It then cleaves off the terminal loop region of the hairpin, about two helical turns away from the Drosha-generated end (Lee et al., 2002; Macrae et al., 2006). This results in a duplex structure ~22 nt long, with 2-nt, 3′ overhangs on both ends. The eventual guide strand would be the miRNA, while the opposing strand is known as the miRNA* (Lau et al., 2001). In conjunction with Dicer, an associated dsRNA binding protein helps mediate processing of the pre-miRNA and the subsequent miRNP/RISC assembly (Chendrimada et al., 2005; Forstemann et al., 2005). The miRNA:miRNA* duplex is then loaded onto Ago[10], after which the miRNA* appears to be peeled away and degraded, or directly cleaved in the case of Ago2. The same thermodynamic rules for selecting guide strands of exogenous siRNAs apply for picking the guide miRNA strand (Khvorova et al., 2003; Schwarz et al., 2003).

With a few exceptions, the vast majority of miRNAs are processed via this pathway. A small group of miRNAs is encoded by introns that are short enough to resemble pre-miRNAs (mirtrons). The pre-miRNA hairpin structure is generated by splicing and subsequent debranching of the mirtron, thus bypassing the Drosha processing step (Berezikov et al., 2007; Okamura et al., 2007; Ruby et al., 2007). Thus far, one miRNA, miR-451, is known to bypass the Dicer processing step. This miRNA does not entirely map to the stem of its pre-miRNA hairpin, as canonical miRNAs do, and its biogenesis is dependent upon Ago2. The miR-451 pre-miRNA is loaded directly onto Ago2, which cleaves its passenger arm. Mature miR-451 is generated after further trimming (Figure 1A)(Cheloufi et al., 2010; Cifuentes et al., 2010).

[10] The loaded Ago:miRNA complex is thought to be extremely stable, with a half-life of hours, or even days (Gatfield et al., 2009; van Rooij et al., 2007). However, there is increasing evidence of regulated miRNA decay. For example, during dark adaptation in mouse retina, several miRNAs are rapidly turned over, with half-lives of about 1 hour (Krol et al., 2010).

MicroRNAs: Target prediction

One of the reasons miRNAs have not been conducive for discovery by genetic approaches is their small size. This gap, however, has been filled by rapid miRNA discovery through high-throughput sequencing.

Equally important to the development of the field has been the prediction of miRNA targets through the use of bioinformatics. In plants, where most miRNAs are perfectly complementary to their targets, it is relative easy to predict miRNA targets computationally. This approach led to the discovery that most targets in plants are transcription factors, contributing to the view that plant miRNAs help clear transcripts in the current developmental stage to pave the way for transitions to daughter cell lineages (Rhoades et al., 2002). Metazoan miRNAs, which tend to have imperfect complementarity to their targets, present a more difficult challenge, and it was only with the use of conservation that bioinformatic approaches began to bear fruit (Enright et al., 2003; Lewis et al., 2003; Stark et al., 2003). When different segments of conserved miRNAs were searched against putative targets for perfect Watson-Crick base-pairing, pairing to nucleotides 2–7 of the miRNA (termed the 'seed') was found to be the most critical (Lewis et al., 2003). The region of the mRNA target that forms Watson-Crick base pairs to the 'seed' was termed the 'seed match' (Lewis et al., 2003). Subsequent studies identified different categories of the seed match (Lewis et al., 2005) and further delineated the hierarchy of target site efficacies (Figure 1B)(Grimson et al., 2007; Nielsen et al., 2007). Unlike siRNAs, miRNAs typically do not specify target cleavage. This, together with the extent of base pairing of some of the earliest miRNA:target interactions identified, led to the belief that the difference in repression modes is primarily due to a bulge at the site of the scissile phosphate during miRNA targeting, which precludes cleavage. However, it has since been determined that pairing to the 3′ region of the miRNA is rare, thus the seed matches specify the vast majority of targeting interactions (Bartel, 2009). The cell, however, does not have the luxury of using the filter of conservation, and there are ten times as many non-conserved targets as there are conserved

targets (Bartel, 2009). This led to efforts to uncover determinants of site efficacy without relying on conservation, culminating in the identification of context features in the 3′UTR[11] that can be used to predict effective target sites (Grimson et al., 2007).

In the last two years, methods that covalently crosslink RNA to Ago have been developed[12]. Coupled with deep-sequencing, this allows rapid identification of miRNA target sites *in vivo* (Chi et al., 2009; Hafner et al., 2010). The sequences found to be most enriched in the Ago-crosslinked sites are precisely the various seed-matched target sites. Whereas in the normal scheme of things, experimental approaches would be used to validate computationally identified trends; in this case, seed match enrichment in the sequenced crosslinked sites was cited as evidence for the success of these biochemical approaches. This underscores the importance of the use of bioinformatics, together with the recognition that conservation reflects biological function, in the early development of the miRNA field.

How might one reconcile seed-based targeting with silencing mediated by endonucleolytic cleavage, which requires almost-perfect complementarity of the target to the small RNA? It is thought that the small RNA is bound to Ago with contacts all along its backbone, such that nucleotides 2–8 are preorganized in a geometry that resembles an A-form helix, to favor Watson-Crick pairing to the target mRNA (Bartel, 2009; Mallory et al., 2004). During seed-based targeting, recognition of a seed-matched site initiates non-cleavage repression; when there is extensive complementarity, the base-paired seed region nucleates further base pairing. This involves a conformational change in Ago, which then places the catalytic triad directly at the cleavage site, whereupon cleavage ensues.

---

[11] Many miRNA target sites can also be found in the coding region, but these are much less efficacious than sites in the 3′UTR (Bartel, 2009). Thus, target prediction algorithms have mainly focused on the latter. It is thought that sites in the 3′UTR are more effective because they are sheltered from the passage of translating ribosomes. This is also thought to be the reason a 'ribosome shadow' exists in the first 15 nt of the 3′UTR, where target sites are less likely to be conserved (Grimson et al., 2007).

[12] More than six years after the bioinformatic identification of the importance of the miRNA 'seed' region.

MicroRNAs: Function

Requiring only a 7-nt match suggests that miRNAs can potentially target a large number of genes, and hence, biological pathways. Indeed, miRNAs have been implicated in the control of numerous biological functions (Bartel, 2004; Bushati and Cohen, 2007), and >60% of mammalian messages are under selective pressure to maintain pairing to miRNAs (Friedman et al., 2009).

This seemed counterintuitive initially, as early examples of miRNA repression in worms suggest that miRNAs act as switches to downregulate a few targets to inconsequential levels for advancement to the next stage of development or the next cell fate (Lee et al., 1993; Moss et al., 1997; Reinhart et al., 2000; Wightman et al., 1993), thus it had been suggested that miRNAs may have evolved to regulate the expression of just a few critical targets (Xiao et al., 2007). However the abundance of conserved targets indicates that evolutionary pressure has been exerted on a multitude of targets for each miRNA, and not just a few critical targets. This phenomenon has been rationalized by the micromanager hypothesis (Bartel and Chen, 2004), in which miRNA targeting interactions can be classified as switch interactions (of which the *lin-4:lin-14* interaction is a classic example), tuning interactions (in which target protein expression has to be kept within a narrow range for optimal function[13]), or neutral interactions (for which the expression range for optimal function is broad enough that miRNA targeting takes place without selective pressure). Each miRNA also has antitargets that are under selective pressure to avoid sites to the miRNA, because the ensuing downregulation would have been too deleterious (Farh et al., 2005; Stark et al., 2005). The switch interactions and tuning interactions thus make up the set of conserved targets, and most miRNA:target interactions are believed to be tuning interactions (Bartel, 2009).

[13] An example is the regulation of activation-induced cytidine deaminase (AID) by miR-155 (Dorsett et al., 2008). AID is essential for initiating class switch recombination and somatic hypermutation in activated B cells to diversify immunoglobulin production. However, its mutagenic activity also creates substrates for chromosomal translocations, which can be especially deleterious when oncogenes are involved. Thus it is essential that AID expression level be tightly regulated for optimal function.

Although multiple miRNA target sites can be present on one 3'UTR, about 90% of conserved targets have only one site to a given miRNA[14] and each miRNA:target interaction typically represses expression by <2-fold (Bartel, 2009). This led to the conundrum of how such subtle changes could be biologically important. However, the fact that so many target sites have been selectively maintained despite the subtle changes indicate that optimal protein expression levels is of utmost importance to animal fitness. This is also manifested by certain miRNA knockouts, in which there are no gross abnormalities under normal conditions but phenotypes emerge under conditions of stress (Li and Carthew, 2005; van Rooij et al., 2007; Xu et al., 2003).

This perhaps also reflects the role of miRNAs in regulatory network buffering (Bartel, 2009; Leung and Sharp, 2010). An example is the role of miR-7 in the *Drosophila* eye (Li et al., 2009). mir-7 is part of a reciprocal negative feedback loop that reinforces the decision for photoreceptor progenitors to differentiate. Flies mutant in miR-7 appear normal, and only exhibit phenotypes with the concurrent introduction of a miR-7 target allele that does not respond to signaling transmitted from the epidermal growth factor receptor (EGFR signaling). Hence, the perturbation of one node (EGFR signaling) in the regulatory network sensitizes the system such that other nodes (like miR-7) are now more important[15]. Thus, miRNA:target interactions can be nodes in elaborate regulatory networks, in which phenotypes would only be observed if more than one node (preferably more) are perturbed. This also suggests that in slightly dampening the protein output of hundreds of genes, miRNAs could be providing a buffer against stress in general (Leung and Sharp, 2010); even without stress, dampening protein output while increasing transcriptional output could lead to more uniform expression levels in cells (Bartel and Chen, 2004), considering that transcription often occurs in bursts (Suter et al., 2011).

---

[14] The 3'UTR of *lin-14*, on the other hand, has three canonical sites to *lin-4*, two of which are within 8–40nt apart, meaning they can act cooperatively (Doench et al., 2003; Grimson et al., 2007; Nielsen et al., 2007). This makes *lin-4*:*lin-14* an outlier in the set of miRNA:target interactions, and also makes *lin-4*:*lin-14* perfectly poised to be discovered by genetic screens.

[15] This is highly reminiscent of the classic enhancer screens in flies (Simon, 1994).

How might these interactions be reflected at the molecular level? What mechanisms mediate the subtle changes in protein output? Before delving into how miRNAs repress their targets, it is prudent to revisit steps along the gene expression pathway with which miRNA repression has been thought to interface.

mRNA stability

In eukaryotes, the stability of an mRNA is inextricably linked to the integrity of its ends, because in general, bulk mRNA turnover takes place by exonucleolytic, rather than endonucleolytic, decay (Garneau et al., 2007). Eukaryotic mRNAs typically possess a 5′ 7-methylguanosine (m^7G) cap (Furuichi et al., 1975; Wei et al., 1975) and a poly(A) tail[16] at the 3′ end (Darnell et al., 1971; Edmonds et al., 1971; Lee et al., 1971). In the cytoplasm, these structures interact with the translation initiation factor eIF4E and the cytoplasmic poly(A)-binding protein (PABP), respectively; these interactions are thought to confer protection to the respective ends of the mRNA (Garneau et al., 2007).

The current model for the turnover of most mRNAs in somatic cells starts with deadenylation. Characterized eukaryotic deadenylases include the CCR4-NOT deadenylase (whose activity is inhibited by PABP), PAN2-PAN3 (whose activity is stimulated by PABP) and PARN (which has cap-dependent activity). In mammalian cells, it has been shown that the PAN2-PAN3 complex carries out initial deadenylation of a β-globin reporter transcript, shortening the poly(A) tail from the usual ~200 nt to ~80 nt, after which the CCR4-NOT complex takes over (Yamashita et al., 2005). Deadenylation generally leads to degradation of the mRNA body (Mangus et al., 2003), though in some cases, it can be reversible[17]. Following deadenylation, the main body of the mRNA can be degraded by one of two irreversible pathways. The 3′→5′ pathway is mediated by the exosome, while in the 5′→3′ pathway,

[16] Exceptions include the replication-dependent histone mRNAs, which have a stem-loop at their 3′ ends, instead of poly(A) tails.

[17] For example, in *Xenopus* oocytes, many maternal mRNAs are stored in a translationally dormant state with short poly(A) tails; during oocyte maturation, cytoplasmic polyadenylation of these mRNAs occurs, which then allows translation to proceed (Richter, 2007).

removal of the m^7G cap is initiated by the decapping enzyme DCP2. This renders the rest of the mRNA vulnerable to the 5′→3′ exonuclease, XRN1. Components of the 5′→3′ decay machinery can often be found in cytoplasmic foci known as P bodies. Although DCP2 and XRN1 are among the most frequently used markers for visualizing P bodies, P body formation is not required for 5′→3′ decay activity, and the relative distribution of these factors between P bodies and the rest of the cytoplasm is unclear (Garneau et al., 2007).

Regulation of mRNA stability

Various *cis*-elements exist that regulate the stability of mRNAs. Although these can be found in the 5′UTR and the coding region, they are most frequently present in the 3′UTR, where interacting protein complexes would be shielded from the passage of ribosomes. One of the most well-studied elements is the AU-rich element (ARE), which often includes one or more copies of the pentamer AUUUA. These elements recruit ARE-binding proteins, such as embryonic lethal abnormal vision (ELAV, also known as HuR) proteins, which in turn modulate the recruitment/activity of the decay machinery. AREs are found in the 3′UTRs of many genes that encode cytokines and proto-oncogenes (Chen and Shyu, 1995) — the inherent instability of these transcripts facilitates rapid changes in abundance when required, such as when immune cells are activated upon antigen exposure.

Other RNA-binding proteins that modulate stability include proteins of the Pumilio family, each of which binds to mRNA subpopulations that encode proteins of related functions in yeast (Gerber et al., 2004). Conversely, stabilizing elements exist on some genes with housekeeping roles. An example is α-globin, which contains cytosine-rich elements in its 3′UTR; these sequences recruit the KH-domain RNA-binding proteins αCP1 and αCP2, which in turn confer stability (Kiledjian et al., 1995).

Eukaryotic translation

Next to the availability/existence of the mRNA, the most important step that contributes towards the making of a functional protein is translation of the mRNA. The process of protein synthesis is highly

energy-consuming (Mathews et al., 2007). It is thus hardly surprising that translation, especially the initiation step of translation, is tightly controlled.

For most mRNAs, initiation is the rate-limiting step of translation (Mathews et al., 2007). Briefly, the small subunit (40S) of the ribosome associates with the initiator tRNA (Met-tRNA$_i$), forming the 43S pre-initiation complex, which is then loaded onto the 5′ end of the mRNA (Figure 2A). From the 5′ end, the 43S complex scans for the first AUG on the mRNA[18], whereupon the 5′-CAU-3′ anticodon of the initiator tRNA recognizes the AUG via complementary base pairing. The formation of this 48S initiation complex then allows the large subunit (60S) to join, forming the fully assembled 80S ribosome that is competent for translation elongation. Each of these steps is mediated by a set of initiation factors, whose presence/absence serve as stimulatory factors or checkpoints along the pathway. As mentioned in the previous section, the 5′ m^{7}G cap is bound by eIF4E. eIF4E is part of the eIF4F complex, which comprises eIF4A, eIF4E and eIF4G. eIF4A is a helicase that mediates ATP-dependent unwinding of secondary structure in the 5′UTR, to facilitate scanning by the 43S complex. The formation of the 43S complex is itself regulated by eIF2, which makes use of GTP hydrolysis to ensure the directionality of its actions. In its GTP-bound form, eIF2 binds Met-tRNA$_i$ to form the eIF2:GTP:Met-tRNA$_i$ ternary complex (eIF2-TC). This step is necessary to bring the initiator tRNA to the 40S subunit. Recognition of the start codon triggers hydrolysis of the GTP associated with eIF2. This then leads to eIF5B-mediated joining of the 60S subunit, with concomitant displacement of eIF2 and other initiation factors from the 40S subunit. eIF5B, itself a ribosome-dependent GTPase, then hydrolyzes its GTP and leaves, thus forming the elongation-competent 80S ribosome (Jackson et al., 2010). Analogous regulatory roles in translation elongation and termination are mediated by elongation factors and release factors respectively.

[18] This is the conventional scanning model of translation initiation in eukaryotes (Kozak, 1978).

Study of translation

Translation is most frequently studied by polysome[19] profiling (Figure 2B). This involves velocity sedimentation through a sucrose gradient to fractionate components in cell lysates. By virtue of their large size, nucleic acid composition, and sheer abundance[20], ribosomes and their subunits can be easily monitored by ultraviolet absorbance at 260 nm. Through this method, single ribosomes can be resolved from polysomes; ribosomal subunits at various stages of initiation can also be distinguished. By making mRNA measurements in the various gradient fractions, it is possible to infer the fraction of an mRNA species that is associated with polysomes, and thus presumably the fraction that is actively translated[21]. Methods to measure mRNA levels in each fraction range from those probing single genes, such as Northern blotting, to the more recent use of microarrays that allows thousands of genes to be probed in one experiment (Melamed and Arava, 2007).

The contacts made between ribosomes, or their individual subunits, with the mRNA are able to obstruct the passage of enzymes such as reverse transcriptases. This attribute is exploited in 'toeprinting' experiments, to probe the 3′ end positions of ribosomes on the mRNA (Figure 2B)(Sachs et al., 2002). An analogous method to probe 5′ end positions is known as 'heelprinting' (Wolin and Walter, 1988). Both methods have been used to study ribosome pauses during translation[22].

---

[19] Polysome, or polyribosome, indicates an mRNA associated with more than one ribosome (Warner et al., 1963).

[20] To put this into perspective, all the mRNAs in the cell constitute only 5% of the RNA population; ribosomal RNA (rRNA) typically make up about 80% of all RNA in the cell, and the rest are mostly tRNAs (Warner, 1999).

[21] After incubating reticulocytes with ^{14}C amino acids to label nascent globin chains, followed by velocity sedimentation of the lysate through a sucrose gradient, most of the radioactive label was found to sediment not with 80S ribosomes (which were therefore deemed translationally 'inactive'), but with structures that sedimented much faster in the gradient (these turned out to be the polysomes)(Warner et al., 1963). Thus association with polysomes is often equated with active translation; and the subpolysomal population, inactive.

[22] In particular, toeprinting was used extensively to tease apart steps along the translation initiation pathway. As various initiation factors join and leave the initiation complexes, their presence and/or the conformational changes they induce result in blockages (of reverse transcriptase) at slightly different positions on the mRNA; this in turn allows the order of events to be mapped (Pestova and Kolupaeva, 2002).

The exact location of a ribosome on an mRNA can also be probed by 'footprinting' experiments, in which mRNA segments occupied by ribosomes are protected from nuclease digest. These protected ribosomal 'footprints' are then identified downstream by direct sequencing of radioactively labeled mRNA, or through probes that base pair to these segments. This footprinting approach was first used to study 70S ribosomes on bacteriophage RNA (Steitz, 1969; Takanami et al., 1965), and eventually paved the way for the confirmation of the Shine-Dalgarno hypothesis[23] (Steitz and Jakes, 1975). Based on the principle of footprinting, ribosome profiling was developed (Ingolia et al., 2009). Instead of using laborious single-gene assays to detect ribosomal footprints, the ribosome-protected fragments are converted into libraries for deep sequencing, thus giving a genome-wide snapshot of where ribosomes are in the cell, at any given point in time (Ingolia et al., 2009).

Regulation of translation

Techniques described in the preceding section have been instrumental in deciphering the many ways in which translation is regulated. Most studied examples involve control at initiation, the rate-limiting step of translation. This is partly because controlling initiation makes the most economic sense, and thus more examples exist for discovery; another reason is that blockages at translation initiation are the easiest to detect by polysome profiling. Because elongation rates for most messages are believed to be similar (Mathews et al., 2007), a block at initiation translates to fewer ribosomes on the message, which is reflected by a shift in the polysome profile of the mRNA population in question.

Translation can be regulated at a gene-specific level or at a global level. A classic example of gene-specific translational control is that mediated by the iron-response element (IRE). The IRE is a

[23] In an elegant experiment, radioactively labeled footprints from the initiator region were added to bacterial ribosomes and allowed to hybridize (Steitz and Jakes, 1975). Colicin E3, a toxin that cleaves 16S rRNA about 50 nt from the 3' end, was then used to liberate a piece of 16S rRNA that is small enough to enter a polyacrylamide gel. Stripping away the ribosomal proteins yields a hybrid rRNA:footprint duplex that is easily distinguishable from the 30-nt footprint alone. The band confirmed the binding of the initiator region of the mRNA to a complementary region in 16S rRNA, the Shine-Dalgarno sequence (Shine and Dalgarno, 1974).

hairpin structure on the mRNA, usually in the 5′UTR[24]. In the ferritin mRNA, this hairpin is located very close to the 5′ cap, and binds the iron response protein (IRP) under iron-deficient conditions, which in turns blocks 43S loading. In iron-replete conditions, iron-bound IRP has reduced affinity for the IRE, and its departure allows the 43S complex to be loaded. The loaded 43S complex scans through the IRE hairpin[25], productive translation ensues, and ferritin protein is made to facilitate the storage of excess iron (Hentze and Kuhn, 1996). Later stages of initiation can also be regulated. An example is the lipoxygenase mRNA, which is translationally repressed during erythroid differentiation, until the late reticulocyte stage. In this case, ten differentiation control elements (DICE) in its 3′UTR recruit hnRNPs K and E1, the binding of which block 60S joining through an as yet unknown mechanism (Ostareck et al., 2001; Ostareck et al., 1997).

Global control of translation typically takes place in times of stress and is most commonly effected by the phosphorylation of eIF2α, a subunit of eIF2. After one round of initiator codon recognition, eIF2 requires the stimulatory activity of eIF2B in order to exchange its GDP for GTP (Figure 2A). When the eIF2α subunit is phosphorylated on its Ser51 residue, eIF2 becomes refractory to stimulation by eIF2B[26]. Because eIF2:GDP has reduced affinity for Met-tRNA$_i$, this results in a shortage of eIF2-TC, leading to a global translation block.

Phosphorylation can be performed by one of four kinases, in response to various forms of stress. These are heme-regulated kinase (HRI, activated during oxidative stress), protein kinase R (PKR, viral infection), GCN2 (amino acid deprivation) and PERK (an excess of unfolded polypeptides in the endoplasmic reticulum [ER]). The last kinase is an integral part of the unfolded protein response (UPR),

[24] In the transferrin receptor mRNA, the IREs are in the 3′UTR, and they control mRNA stability instead.

[25] The small ribosomal subunit is capable of scanning through secondary structure in the 5′UTR, as long as it is below a threshold level (about –60 kcal/mol) of thermodynamic stability (Kozak, 1991; Pelletier and Sonenberg, 1985); the IRE hairpin is about –7 kcal/mol (Hentze and Kuhn, 1996), and thus presents no difficulty to scanning.

[26] eIF2α-phosphorylated eIF2 cannot exchange its GDP, but can still bind eIF2B. The amount of eIF2B in cells is limiting compared to eIF2, thus eIF2B can be rapidly sequestered even if only a fraction of the cell's eIF2α were phosphorylated, leading to a swift block in translation (Mathews et al., 2007).

during which only a select population of messages in the cell is translated, while the majority are translationally shut down. During the UPR, part of the cell's arsenal to prevent overwhelming the ER's folding capacity is to block translation globally by phosphorylating eIF2α and reducing eIF2-TC formation. At the same time, a stress-specific transcriptional program is initiated to help overcome the stressful period (Todd et al., 2008). One of the transcription factors involved is ATF4. ATF4 has two upstream open reading frames (uORFs), one of which (uORF2) overlaps, and is out of frame with, the main ORF. uORF1 is constitutively translated and the 40S subunit continues to scan along the 5′UTR after translating uORF1[27]. When eIF2-TC is abundant, the scanning 40S acquires another eIF2-TC in time to translate uORF2; when eIF2-TC is depleted under stressful conditions, there is a higher probability that the initiation codon of uORF2 will be missed while that of the main ORF will be recognized[28]. eIF2-TC depletion thus paradoxically enables ATF4 to be translated, to mediate coordinate transcriptional changes in the UPR (Lu et al., 2004; Vattem and Wek, 2004).

In general, however, uORFs are thought to interfere with translation of the main ORF (Morris and Geballe, 2000). uORFs that initiate with an AUG start codon are present in ~50% of mammalian genes and they correlate with reduced protein expression from the downstream ORF (Calvo et al., 2009). Recently the inventory of uORFs was greatly expanded when ribosome profiling found evidence of widespread uORF translation in yeast (Ingolia et al., 2009). In addition to canonical AUG-initiated uORFs, numerous uORFs initiated by non-canonical start codons were documented, indicating that uORF translation from non-canonical initiation codons is a more widespread phenomenon than previously appreciated. It remains to be seen whether such widespread uORF translation contributes extensively to translational regulation of the main ORF.

---

[27] Resumption of scanning and reinitiation by ribosomes usually occurs only if the first ORF is short and does not have secondary structure strong enough to cause pausing (Kozak, 2001). uORF1 of ATF4, being only three amino acids long, is thus permissive for scanning resumption.

[28] Here, the larger distance between uORF1 and the start codon of the main ORF translates (pun not intended) directly to an increased amount of time available for eIF2-TC acquisition to occur. Yeast have a similar gene, GCN4, that plays an analogous role but with four uORFs, instead of two (Hinnebusch, 2005).

Besides eIF2, phosphorylation events also regulate eIF4E. In the eIF4F cap-binding complex, the interaction between eIF4E and eIF4G can be inhibited by a family of eIF4E-binding proteins (4E-BPs). 4E-BPs, when hypophosphorylated, bind strongly to eIF4E, preventing the latter's interaction with eIF4G. This compromises cap-dependent translation, which is stimulated by the interaction between eIF4E and eIF4G (Figure 2A; see later section). Thus, 4E-BP phosphorylation is usually induced in times of growth, when translation needs to be upregulated globally (Gingras et al., 1999). Gene-specific forms of 4E-BPs also exist. One example is Cup, an eIF4E-binding protein that is recruited to *nanos* mRNA by Smaug (Nelson et al., 2004) and to *oskar* mRNA by Bruno (Nakamura et al., 2004). Smaug and Bruno, in turn, are recruited by corresponding elements in the respective 3′UTRs. This mechanism helps mediate spatial translational repression of *oskar* and *nanos* mRNAs, and is essential to ensure proper development of the *Drosophila* embryo (Vardy and Orr-Weaver, 2007).

Interestingly, when translation initiation is blocked, translationally repressed mRNAs have been observed to accumulate in stress granules (Anderson and Kedersha, 2008). Stress granules and P bodies differ in that the mRNAs in stress granules are associated with 40S subunits and a select group of initiation factors, both of which are not found in P bodies. Stress granules also do not contain mRNA decay enzymes, unlike P bodies; it is thus thought that mRNAs in stress granules are perhaps waiting to be triaged — to P bodies for degradation, stored as translationally silent mRNAs, or back to the translating pool (Anderson and Kedersha, 2008).

Certain mRNAs, especially those of viruses, initiate translation by a cap-independent mechanism. This involves the use of internal ribosome entry sites (IRESs). IRESs are long, highly structured sequence elements in 5′UTRs that are capable of recruiting 40S subunits directly, bypassing the need for a 5′ cap (Hellen and Sarnow, 2001). Naturally, this obviates the need for eIF4E, but different IRESs have different requirements, and some dispense with the eIF4F complex entirely. An example is the IRES of

hepatitis C virus (HCV), which only requires eIF2 and eIF3 for initiation, and if eIF2-TC is in short supply, it could also recruit Met-tRNA$_i$ with just the use of eIF5B[29] (Terenin et al., 2008).

Global control of translation can also be effected by regulating the synthesis of ribosomes. In mammals, mRNAs encoding ribosomal proteins typically begin with m^{7}GpppC, followed by a run of pyrimidines (Meyuhas, 2000). This is known as the 5′-terminal oligopyrimidine tract (5′TOP). The 5′TOP motif confers coordinate translational regulation on the ribosomal proteins and some translation factors; 5′TOP mRNAs are translationally repressed in quiescent cells, and strongly upregulated during growth, though the exact mechanim by which this occurs is still unclear.

Interplay between translation and mRNA stability

Translation and mRNA stability are thought to be intimately linked, and several mRNA quality control pathways are dependent upon translation. A classic example is the nonsense-mediated decay (NMD) pathway, in which mRNAs possessing a premature termination codon (PTC) are rapidly degraded. While there are slight mechanistic differences between the NMD pathways in yeast and metazoans, both are triggered by a PTC in poor context for translation termination, which leads to the stalling of terminating ribosomes (Garneau et al., 2007; Isken and Maquat, 2007). In yeast, non-stop decay (NSD) and no-go decay (NGD) target mRNAs without stop codons and mRNAs with strong secondary structure in the coding region respectively; both scenarios lead to stalled ribosomes[30]. Both NSD and NGD rely on proteins that mimick translation factors that are normally involved in mediating elongation and termination. This mimicry imbues the proteins with the ability to interact with the A site of the ribosome,

[29] eIF5B is the eukaryotic homolog of bacterial IF2. In bacteria, IF2 is the initiation factor that brings in the initiator tRNA. The ability to recruit the initiator tRNA using eIF5B bypasses the need for eIF2. This can be particularly advantageous to the virus when eIF2 is inactivated by PKR phosphorylation of eIF2α in virus-infected cells, with the concomitant global shortage of eIF2-TC.

[30] Strong secondary structure in the coding region stalls the ribosome in NGD; in NSD, the ribosome translates all the way through the poly(A) tail, eventually stalling at the end of the tail.

thus allowing disassembly of an unproductive ribosome before triggering degradation of the transcript (Doma and Parker, 2007). Interestingly, in yeast, the ability of ribosomes to carry out translation is itself proofread by the cell. Translationally defective rRNA is eliminated by the nonfunctional rRNA decay (NRD) pathway (LaRiviere et al., 2006). Defective 18S and 25S rRNAs are detected by different mechanisms; intriguingly, the 18S NRD pathway makes use of the same detection system as NGD, except that the ribosome stalls due to the intrinsic rRNA defect[31], rather than due to unpassable secondary structure in the mRNA (Cole et al., 2009).

mRNA stability is also thought to be positively correlated with translation efficiency. For example, the disruption of certain translation initiation factors has been shown to decrease mRNA stability (Schwartz and Parker, 1999). At first glance, it may seem like the act of translation protects the mRNA from degradation[32]. However, this interpretation can be complicated by the fact that the factors disrupted in this study are the same factors present in the cap-binding complex, which, by its interaction with PABP, could play a role in protecting the integrity of both ends of the mRNA.

Interaction between the 5′ and 3′ ends of an mRNA

Early experiments found that adding exogenous poly(A) RNA *in vitro* inhibited the translation of poly(A)$^+$ mRNA (Grossi de Sa et al., 1988; Jacobson and Favreau, 1983), which could be reversed by adding purified PABP (Grossi de Sa et al., 1988). A subsequent comparison of poly(A)$^+$ and poly(A)$^-$ reporter mRNAs found that the poly(A) tail conferred a two to three fold stimulation of translation that was not attributable to changes in mRNA stability (Munroe and Jacobson, 1990), suggesting that the

[31] Defects in the 18S and 25S rRNAs do not prevent ribosomal subunit assembly. Hence in the case of defective 18S rRNA, defective small subunits are still assembled and can still translate, until the intrinsic defect in tRNA proofreading finally stalls the ribosome and triggers NRD. In contrast, large subunits with defective 25S rRNA are not associated with monosomes or polysomes, and do not enter the translating pool.

[32] In fact, in bacteria, in which mRNAs do not have 5′ caps or poly(A) tails, this has been shown to be the case (Fan et al., 1964).

poly(A) tail and PABP could play a role in stimulating translation. mRNAs with both a 5′ cap and a 3′ poly(A) tail also exhibited a synergistic enhancement of translation that was not seen in messages missing one component or the other, further suggesting that the two ends could physically interact to stimulate translation (Gallie, 1991). This emerging idea dovetailed with the observation of circular polysomes on the rough ER of rat pituitary cells (Christensen et al., 1987).

Direct physical interactions between eIF4G, eIF4E (which binds the 5′ cap) and PABP (which binds the poly(A) tail) is thought to bring about a 'closed loop' conformation that is capable of stimulating translation (Figure 2A)(Kahvejian et al., 2001). In support of this model, pseudo-circular structures were observed by atomic force microscopy when recombinant yeast eIF4G, PABP, eIF4E and an RNA template possessing a 5′ cap and 3′ poly(A) tail were mixed *in vitro* (Wells et al., 1998). The exact mechanism by which circularization stimulates translation is not known, though it is thought that circularization may promote the recruitment of ribosomal subunits (Sachs et al., 1997). Interestingly, replication-dependent histone mRNAs, which are not polyadenylated and instead have a histone stem loop at the 3′ end, are also circularized. This is mediated by indirect interactions between the stem-loop binding protein (SLBP) and eIF4G, with SLIP1 bridging the two; this interaction is required for efficient translation of histone mRNAs (Cakmakci et al., 2008). Similarly, rotavirus mRNAs, which are capped but not polyadenylated, are also circularized. In this case, a conserved sequence in the 3′UTR is recognized by the viral protein NSP3, which in turn interacts with eIF4G[33] to stimulate translation (Vende et al., 2000). These findings speak to the biological significance of circularization towards stimulating translation. It is thought that promoting translation by circularization helps ensure that only intact mRNAs can act as efficient templates for translation (Kahvejian et al., 2001).

[33] Furthermore, NSP3 interacts with the same region of eIF4G as PABP, thus it competitively displaces PABP and confers further advantage to the virus by simultaneously interfering with efficient translation of host mRNAs which are polyadenylated (Piron et al., 1998).

Molecular consequences of microRNA-mediated repression

Where do miRNAs come in against the myriad mechanisms of post-transcriptional regulation? Most plant miRNAs have extensive complementarity to their target mRNAs, which allows them to specify endonucleolytic cleavage (Jones-Rhoades et al., 2006). Although such miRNA:target interactions do exist in animals (Davis et al., 2005; Yekta et al., 2004), they are extremely rare. Thus, the difficulty posed by most animal miRNAs to computational biologists developing target prediction methods also presented a challenge to wet-lab biologists trying to decipher the mechanism of repression. Many of the methods described in the previous sections have been used to dissect this question, and many of the known post-transcriptional regulatory pathways have been proposed as mechanisms by which miRNAs repress their targets.

Just as *lin-4* had shaped the thinking in the field in terms of miRNA:target interactions, it has had a similar effect on the mechanistic branch of the field. The observation that *lin-4* represses LIN-14 protein output, without substantially changing *lin-14* mRNA levels (Wightman et al., 1993) prompted the use of polysome profiling to pinpoint the stage of translation that was blocked. The first such study reported that the *lin-14* mRNA was found primarily in heavy polysomes and that its distribution across the different fractions was not changed by *lin-4* regulation (Olsen and Ambros, 1999). A similar observation was made with *lin-28*, another target of *lin-4* (Seggerson et al., 2002). Because the polysome profiles did not change with *lin-4* repression, it was suggested that translation was blocked at a post-initiation step.

Conversely, other studies reported a marked shift in sedimentation towards the top of the gradient when a reporter target RNA contained functional miRNA target sites, or when Ago2 was tethered to the reporter (Bhattacharyya et al., 2006; Pillai et al., 2005), suggesting that ribosome loading was reduced with miRNA repression, and implying a block at translation initiation. Various mechanisms have been

proposed to explain the reduction in ribosome loading; these include the impairment of eIF4E function (Humphreys et al., 2005), inhibition of 60S joining (Chendrimada et al., 2007) and localization to P bodies, which are devoid of ribosomes (Bhattacharyya et al., 2006). Yet other studies came to the conclusion that miRNAs inhibit translation at a post-initiation stage — consistent with the two initial reports — and proposed premature ribosome drop-off (Petersen et al., 2006), elongation slow-down (Maroney et al., 2006), and cotranslational degradation (Nottrott et al., 2006; Olsen and Ambros, 1999) as possible mechanisms of post-initiation blockage. In addition to polysome profiling, reporter constructs containing IRESs were often used to determine whether miRNAs block cap-dependent translation, but such studies often came to opposite conclusions. In some studies, the IRES constructs were refractory to miRNA-mediated repression suggesting miRNAs target some step during cap-dependent initiation (Humphreys et al., 2005; Pillai et al., 2005); in others, the IRES constructs were still repressed, suggesting a post-initiation block (Lytle et al., 2007; Petersen et al., 2006). Attempts at *in vitro* reconstitution of miRNA-mediated repression presented a more uniform view. Reconstitution experiments have implicated translation initiation as the step repressed by miRNAs (Mathonnet et al., 2007; Thermann and Hentze, 2007; Wakiyama et al., 2007; Wang et al., 2008). In particular, one study found that the addition of purified eIF4F was able to reduce the extent of miRNA-mediated repression (Mathonnet et al., 2007), which is in agreement with the previously reported involvement of eIF4E (Humphreys et al., 2005). An important caveat must be considered for many of these studies. Because assays measuring final protein output are often used in these studies, any repressive effect of miRNAs can only be seen if miRNA targeting affects the rate-limiting step of translation (Nissan and Parker, 2008) — for example, if miRNAs cause the translation of targets to slow down two-fold at a step that is normally three times as fast as the rate-limiting step, this "repression" will be invisible to methods that rely on readouts that measure the final output of functional protein[34], as that will be limited by the rate of the slowest step.

[34] Such as luciferase assays.

It is also worth noting that one piece of evidence often cited for miRNAs mediating post-initiation translational repression is that miRNAs are often found to cosediment with polysomes (Kim et al., 2004; Maroney et al., 2006; Nottrott et al., 2006). In hindsight, this is an indication of the influence that the *lin-4*:*lin-14* interaction had on the development of the field: at the time, it was not clear that a single miRNA:target site interaction typically mediates repression of <2-fold, thus the fact that *lin-4* repressed *lin-14* so strongly led to the thinking that if a target were repressed at the initiation stage, the almost-complete translational arrest must mean that miRNAs would not be associated with polysomes, and the converse must mean that translational repression is occurring at a post-initiation step.

Even as controversy rages around the mode of translation repression, a similar debate over the extent to which translational repression contributes towards the overall repression was also ongoing. The early paradigm of miRNAs only mediating translational repression was challenged when it was found that target mRNA levels were also reduced for the exact same miRNA:target pair, *lin-4*:*lin-14*, that was used in the original studies (Bagga et al., 2005)[35]. In addition, transfection of miRNA duplexes into HeLa cells led to reduced mRNA levels for hundreds of messages with sequences complementary to the seed of the transfected miRNA (Lim et al., 2005); the converse was observed when an antagomir was introduced to compromise miR-122 repression (Krutzfeldt et al., 2005). Subsequent work showed that the reduction in target mRNA levels mediated by miRNAs is not due to endonucleolytic cleavage but involves deadenylation (Behm-Ansmant et al., 2006; Giraldez et al., 2006; Wu et al., 2006). Components in both pathways of the normal mRNA turnover machinery, including the CCR4-NOT deadenylase complex and the DCP1:DCP2 decapping complex, have since been implicated in target mRNA destabilization (Behm-Ansmant et al., 2006; Rehwinkel et al., 2005). Hence, although the mRNA destabilization effect was

[35] It is thought that the perceived differences in mRNA changes could be due to the different measurement methods used. The original studies had used ribonuclease protection assays (RPAs), which are not able to detect any loss of integrity of the mRNA, unlike a Northern (used in Bagga et al., 2005), which allows full length mRNA to be distinguished from partially degraded pieces. This observed discrepancy also led to the current perception that upon miRNA targeting, mRNA decay takes place slowly; because if decay were rapid, then one might not expect such a difference between using RPAs or Northern blotting.

discovered later, an increasing body of evidence indicates that it does take place and the mechanism by which it occurs is more understood than that of miRNA-mediated translational repression.

Together with the observation that Ago proteins and targeted mRNAs can be localized to P bodies (Liu et al., 2005; Pillai et al., 2005; Sen and Blau, 2005), the involvement of the decapping enzymes led to the idea that P bodies could play a role in miRNA-mediated repression. P body involvement was also appealing because in one study, translationally repressed miR-122-targeted CAT-1 mRNA was found to localize to P bodies, and the repression could be reversed upon stress, as manifested by reassociation with polysomes[36] (Battacharryya et al, 2006). Although this study also noted that CAT-1 mRNA did not localize to stress granules upon arsenite treatment, this is not unexpected as endogenous CAT-1 is translated from an IRES in its 5'UTR upon stress induced by eIF2α phosphorylation (Fernandez et al., 2002), and arsenite treatment triggers eIF2α phosphorylation via the HRI kinase; thus it would be unlikely for CAT-1 to localize to stress granules — which consists of translationally repressed mRNAs — in the first place. Interestingly, a quantitative fluorescence study found that a subpopulation of Ago proteins, miRNAs and mRNA targets become enriched in stress granules upon various stresses in a miRNA-dependent manner (Leung et al., 2006). Thus far, however, the role of P bodies and stress granules in miRNA target repression is still unclear.

One common P body component that does play an important role in miRNA-mediated repression is GW182 (Figure 1A), though its localization to P bodies is not necessary for repression (Eulalio et al., 2009a; Lazzaretti et al., 2009). GW182 proteins interact with Ago proteins through Gly-Trp (GW) repeats. They are essential for miRNA-mediated repression, and likely act downstream of Ago (Eulalio et

[36] Translational repression is relieved when the HuR protein, which is released from the nucleus during stress, binds to AREs in the CAT-1 3'UTR and compromises miR-122 repression (Bhattacharyya et al., 2006). A similar interplay between the miRNP and other 3'UTR binding proteins occurs in zebrafish: some miR-430 targets, such as *nanos1* and *tdrd7*, are repressed in somatic cells but not in the germline because the binding of germline-expressed Dnd1 to nearby U-rich regions in the 3'UTR prevents binding of the miRNP (Kedde et al., 2007).

al., 2009c). Recently, *Drosophila* GW182 and human TNRC6C[37] were reported to interact with cytoplasmic PABP (Fabian et al., 2009; Jinek et al.; Zekri et al., 2009), and this contact was required for maximal miRNA-mediated deadenylation. In particular, GW182 could compete with eIF4G for binding to PABP, suggesting that disrupting the 'closed loop' conformation mediated by the eIF4E:eIF4G:PABP interaction could be the mechanism by which miRNAs repress their targets. In addition, the GW182:PABP interaction was reported to facilitate the recruitment of the CCR4-NOT deadenylase complex to the target mRNA. Interestingly, a subsequent study showed that a β-globin reporter with miRNA target sites is destabilized in a biphasic manner, first by the PAN2-PAN3 deadenylase complex, then by the CCR4-NOT complex (Chen et al., 2009); this same pathway was previously reported for normal mRNA turnover (Yamashita et al., 2005), just that in the presence of miRNA targeting, the kinetics are accelerated. Hence, not only could breaking the 'closed loop' conformation expose the ends of the targeted mRNA to the decay machinery, there is likely to be direct stimulation of mRNA destabilization by recruitment of the decay enzymes. Concurrently, the synergistic enhancement of translation afforded by the 'closed loop' would be lost, likely resulting in reduced ribosome loading, which would be reflected as translational repression (Huntzinger and Izaurralde, 2011).

These findings would predict that having a poly(A) tail could enhance the magnitude of repression, which is indeed consistent with some of the earlier studies that had observed greater repression with polyadenylated reporter mRNAs (Humphreys et al., 2005; Iwasaki et al., 2009; Wakiyama et al., 2007; Wu et al., 2006). In particular, in one study that compared repression between a polyadenylated construct and a non-polyadenylated one that ends with a histone stem loop, both reporters were repressed to a similar extent in terms of translational efficiency, but the polyadenylated reporter was repressed more due to changes that could be attributed to mRNA destabilization (Wu et al., 2006). However, similar studies in which more reporters were interrogated have reported inconsistent results (Eulalio et al., 2008; Eulalio et al., 2009b), thus this hypothesis remains to be rigorously tested.

[37] There are three paralogs of GW182 in humans — TNRC6A, TNRC6B and TNRC6C — versus one in *Drosophila*.

To maximize the probability of observing repression in many of these studies, the reporter constructs used often have multiple sites to the same miRNA, which is way more than the one or two sites that endogenous targets typically have (Bartel, 2009). Even when endogenous 3′UTRs were used in reporter constructs, they often present inconsistent views with regards to the relative contributions from translational repression and mRNA destabilization. It was thus of interest to learn to what extent each component contributes towards the overall miRNA-mediated repression, especially at a genome-wide level. Initial large-scale studies were typically done with microarrays, which would miss targets that were only translationally repressed. To address this void, quantitative proteomics approaches, with simultaneous mRNA measurements by microarray, were used to monitor protein levels directly (Baek et al., 2008; Selbach et al., 2008). Reassuringly, these proteomics experiments observed the same hierarchy of target site efficacies (Figure 1B) as delineated by previous microarray studies (Grimson et al., 2007; Nielsen et al., 2007), indicating that the contribution from mRNA destabilization effects was not insubstantial compared to the overall repression of protein output. Proteomics approaches that measure steady-state protein levels, though, have to contend with the potentially confounding factor of protein degradation rates. Such approaches can also be biased towards detecting highly abundant proteins, and have a smaller dynamic range than genomics-based approaches. Hence, more recently, the humble polysome profiling method was modified with a gradient-encoding technique to look at translation directly and bypass the need to measure protein levels (Hendrickson et al., 2009).

It is against this backdrop that research described in this thesis is conducted, and furious action is still taking place around the world to unravel the manner by which miRNAs mediate repression. While the basic tenet of miRNA targeting looks extremely similar to Model II in Jacob and Monod's review article, the details have turned out to be slightly more complicated. However, compared to other examples of post-transcriptional regulation that are mediated by 3′UTR-protein interactions, miRNA targeting is, evolutionarily, a simpler way of tweaking protein output. As opposed to protein-based recognition of 3′UTR sequence determinants, using seed-based targeting to harness existing machinery

(Ago, GW182, and other yet-to-be-identified players) is precisely the reason it is possible to make the incremental changes required to sample expression space for optimizing protein output. This underscores just how fascinating and multi-faceted biology can be. An additional bright side is: it can safely be said that the field of post-transcriptional control has come into its own and is no longer the poorer cousin of transcriptional control.

Figure legends

Figure 1. The biogenesis of microRNAs and their targeting specificity.

(A) Biogenesis pathways of miRNAs. The canonical miRNA biogenesis pathway is represented by a monocistronic miRNA locus. Processing factors are labeled with the names of their human orthologs. In the canonical biogenesis pathway, the pri-miRNA is processed by the Microprocessor complex (Drosha, with its cofactor: DGCR8 in mammals; Pasha in *Drosophila* and *C. elegans*) to liberate the pre-miRNA — a process known as 'cropping'. The pre-miRNA, with its base-paired stem and short 3′ overhang, is recognized by Exportin-5 and exported to the cytoplasm. In the cytoplasm, Dicer recognizes the pre-miRNA with the help of an associated dsRNA binding protein (TRBP in mammals; Loquacious in flies) and catalyzes the second processing step ('dicing') to yield a miRNA:miRNA* duplex with 2-nt 3′ overhangs. The duplex is loaded onto Ago, whereupon only the guide strand is kept. The assembled miRNP complex then directs targeting. Note that the miRNP is represented minimally by Ago and the miRNA, but actually consists of additional components. If pairing to the target mRNA is extensive, Ago2-mediated endonucleolytic cleavage ensues. Typically, pairing is only mediated through the miRNA 'seed' region. This directs repression by translational repression and/or mRNA destabilization via the bulk mRNA turnover pathways. In addition to Ago, GW182 is an integral player in directing this type of repression.

Mirtrons are embedded in short introns that upon splicing and debranching, yield the pre-miRNAs, thus bypassing the Drosha processing step. The pre-miRNA for miR-451 is loaded directly onto Ago2, which cleaves its passenger arm; further trimming of the passenger arm results in an assembled miR-451 miRNP, thus bypassing the Dicer processing step.

(B) Different types of miRNA target sites. Shown are miRNA target sites, according to the established hierarchy of targeting efficacies. The seed region (highlighted in orange) of the miRNA determines targeting specificity. Base-paired nucleotides opposite the seed (highlighted in blue) are part of the 'seed match'. The seed match can be made more efficacious by a Watson-Crick base pair at position 8 of the miRNA (7mer-m8), or an adenosine moiety across from position 1 (7mer-A1), or both (8mer). Bioinfomatic analyses and experimental evidence have delineated the hierarchy of target site efficacies as indicated (Friedman et al., 2009; Grimson et al., 2007; Nielsen et al., 2007).

Figure 2. Translation initiation and the methods used to study translation.

(A) Current model of eukaryotic translation initiation pathway. Translation initiation factors are labeled without the 'eIF' prefix for clarity. Following termination, the 40S subunit is prevented from associating with the 60S subunit by the binding of eIF3, eIF1 and eIF1A. The 43S pre-initiation complex (43S PIC) is formed with the joining of the eIF2:GTP:Met-tRNA$_i$ ternary complex (eIF2-TC). mRNA activation occurs with the binding of the eIF4F cap binding complex and eIF4B. eIF4F binding brings about circularization of the mRNA via the interaction between PABP (which is bound to the poly(A) tail), eIF4G, and eIF4E (which binds to the 5′ cap). This 'closed loop' conformation is thought to stimulate translation initiation. The binding of eIF4F also unwinds the cap-proximal region of the mRNA to prepare it for ribosome attachment. The 43S PIC is loaded near the cap with the help of the eIF4F complex, and scanning for the AUG start codon ensues. Scanning requires ATP and is mediated by the eIF4A helicase, whose activity is stimulated by eIF4B. Recognition of the initiation codon triggers GTP

hydrolysis by eIF2; this GTPase activity is stimulated by eIF5. eIF2's affinity for Met-tRNA$_i$ and the 40S subunit is reduced after GTP hydrolysis, facilitating its displacement when the 60S subunit joins. eIF5B mediates 60S joining, with the concomitant release of many of the initiation factors associated with the 48S initiation complex. The 80S ribosome thus assembled is only competent for elongation after eIF5B, a ribosome-dependent GTPase, hydrolyzes its GTP and leaves. The recycled eIF2:GDP exchanges its GDP for GTP, with the help of eIF2B. This step is necessary before eIF2 is able to bind Met-tRNA$_i$ to form the ternary complex.

(B) Methods used to study translation. A polysome profile generated by velocity sedimentation through a sucrose gradient can distinguish ribosomal subunits from 80S ribosomes and polysomes as shown. RNA extracted from each fraction can then be probed to assess the proportion of a given mRNA species that is associated with polysomes. Depending on the gradient used and the time/speed of sedimentation, this method can also distinguish between different stages of initiation, such as the 43S and 48S complexes (not shown).

Ribosome footprints generated from nuclease digest give the locations of ribosomes on an mRNA by sequence identification. In 'toeprinting', a radioactively labeled primer (labeled at the 5′ end, denoted by a red asterisk) is extended by reverse transcriptase; the enzyme is blocked when it hits the 3′ end of the ribosome, and the accumulation of such blocks generates a strong band at a discrete position when the primer-extended products are run on a gel. This method maps the 3′ ends of paused ribosomes. In 'heelprinting', ribosome footprints are hybridized to an antisense complementary DNA (cDNA). T4 DNA polymerase is used to extend an end-labeled primer; because the polymerase cannot unwind the RNA:DNA hybrid when it hits the footprint, it stalls and the accumulation of such blocks again results in a strong band at a discrete position when the primer-extended products are run on a gel. This method maps the 5′ ends of paused ribosomes. For clarity, only a single ribosome is shown for toeprinting, and similarly, only a single footprint is shown for heelprinting; note that the directional nature of these assays

mean that if there are multiple translational pauses on an mRNA, the readout will be biased towards the first pause encountered.

References

Abel, P.P., Nelson, R.S., De, B., Hoffmann, N., Rogers, S.G., Fraley, R.T., and Beachy, R.N. (1986). Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene. Science *232*, 738-743.

Ambros, V. (1989). A hierarchy of regulatory genes controls a larva-to-adult developmental switch in C. elegans. Cell *57*, 49-57.

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in C. elegans. Curr Biol *13*, 807-818.

Anandalakshmi, R., Pruss, G.J., Ge, X., Marathe, R., Mallory, A.C., Smith, T.H., and Vance, V.B. (1998). A viral suppressor of gene silencing in plants. Proc Natl Acad Sci U S A *95*, 13079-13084.

Anderson, P., and Kedersha, N. (2008). Stress granules: the Tao of RNA triage. Trends Biochem Sci *33*, 141-150.

Angell, S.M., and Baulcombe, D.C. (1997). Consistent gene silencing in transgenic plants expressing a replicating potato virus X RNA. EMBO J *16*, 3675-3684.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., *et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. Dev Cell *5*, 337-350.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev *22*, 2773-2785.

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature *455*, 64-71.

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A.E. (2005). Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. Cell *122*, 553-563.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Bartel, D.P., and Chen, C.Z. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet *5*, 396-400.

Beclin, C., Berthome, R., Palauqui, J.C., Tepfer, M., and Vaucheret, H. (1998). Infection of tobacco or Arabidopsis plants by CMV counteracts systemic post-transcriptional silencing of nonviral (trans)genes. Virology *252*, 313-317.

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes Dev *20*, 1885-1898.

Berezikov, E., Chung, W.J., Willis, J., Cuppen, E., and Lai, E.C. (2007). Mammalian mirtron genes. Mol Cell *28*, 328-336.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Bhattacharyya, S.N., Habermacher, R., Martine, U., Closs, E.I., and Filipowicz, W. (2006). Relief of microRNA-mediated translational repression in human cells subjected to stress. Cell *125*, 1111-1124.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Brenner, S., Jacob, F., and Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature *190*, 576-581.

Brigneti, G., Voinnet, O., Li, W.X., Ji, L.H., Ding, S.W., and Baulcombe, D.C. (1998). Viral pathogenicity determinants are suppressors of transgene silencing in Nicotiana benthamiana. EMBO J *17*, 6739-6746.

Bushati, N., and Cohen, S.M. (2007). microRNA functions. Annu Rev Cell Dev Biol *23*, 175-205.

Cakmakci, N.G., Lerner, R.S., Wagner, E.J., Zheng, L., and Marzluff, W.F. (2008). SLIP1, a factor required for activation of histone mRNA translation by the stem-loop binding protein. Mol Cell Biol *28*, 1182-1194.

Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci U S A *106*, 7507-7512.

Cerutti, H., and Casas-Mollano, J.A. (2006). On the origin and functions of RNA-mediated silencing: from protists to man. Curr Genet *50*, 81-99.

Chalfie, M., Horvitz, H.R., and Sulston, J.E. (1981). Mutations that lead to reiterations in the cell lineages of C. elegans. Cell *24*, 59-69.

Cheloufi, S., Dos Santos, C.O., Chong, M.M., and Hannon, G.J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature *465*, 584-589.

Chen, C.Y., and Shyu, A.B. (1995). AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem Sci *20*, 465-470.

Chen, C.Y., Zheng, D., Xia, Z., and Shyu, A.B. (2009). Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. Nat Struct Mol Biol *16*, 1160-1166.

Chendrimada, T.P., Finn, K.J., Ji, X., Baillat, D., Gregory, R.I., Liebhaber, S.A., Pasquinelli, A.E., and Shiekhattar, R. (2007). MicroRNA silencing through RISC recruitment of eIF6. Nature *447*, 823-828.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436*, 740-744.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479-486.

Christensen, A.K., Kahn, L.E., and Bourne, C.M. (1987). Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary. Am J Anat *178*, 1-10.

Cifuentes, D., Xue, H., Taylor, D.W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G.J., Lawson, N.D., *et al.* (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science *328*, 1694-1698.

Cogoni, C., Irelan, J.T., Schumacher, M., Schmidhauser, T.J., Selker, E.U., and Macino, G. (1996). Transgene silencing of the al-1 gene in vegetative cells of Neurospora is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. EMBO J *15*, 3153-3163.

Cogoni, C., and Macino, G. (2000). Post-transcriptional gene silencing across kingdoms. Curr Opin Genet Dev *10*, 638-643.

Cole, S.E., LaRiviere, F.J., Merrikh, C.N., and Moore, M.J. (2009). A convergence of rRNA and mRNA quality control pathways revealed by mechanistic analysis of nonfunctional rRNA decay. Mol Cell *34*, 440-450.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., *et al.* (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798-802.

Darnell, J.E., Wall, R., and Tushinski, R.J. (1971). An adenylic acid-rich sequence in messenger RNA of HeLa cells and its possible relationship to reiterated sites in DNA. Proc Natl Acad Sci U S A *68*, 1321-1325.

Davis, E., Caiment, F., Tordoir, X., Cavaille, J., Ferguson-Smith, A., Cockett, N., Georges, M., and Charlier, C. (2005). RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. Curr Biol *15*, 743-749.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231-235.

Doench, J.G., Petersen, C.P., and Sharp, P.A. (2003). siRNAs can function as miRNAs. Genes Dev *17*, 438-442.

Doma, M.K., and Parker, R. (2007). RNA quality control in eukaryotes. Cell *131*, 660-668.

Dorsett, Y., McBride, K.M., Jankovic, M., Gazumyan, A., Thai, T.H., Robbiani, D.F., Di Virgilio, M., Reina San-Martin, B., Heidkamp, G., Schwickert, T.A., *et al.* (2008). MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. Immunity *28*, 630-638.

Edmonds, M., Vaughan, M.H., Jr., and Nakazato, H. (1971). Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship. Proc Natl Acad Sci U S A *68*, 1336-1340.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001a). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev *15*, 188-200.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001b). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J *20*, 6877-6888.

Eulalio, A., Helms, S., Fritzsch, C., Fauser, M., and Izaurralde, E. (2009a). A C-terminal silencing domain in GW182 is essential for miRNA function. RNA *15*, 1067-1077.

Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008). GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. Nat Struct Mol Biol *15*, 346-353.

Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., and Izaurralde, E. (2009b). Deadenylation is a widespread effect of miRNA regulation. RNA *15*, 21-32.

Eulalio, A., Tritschler, F., and Izaurralde, E. (2009c). The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. RNA *15*, 1433-1442.

Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A., *et al.* (2009). Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. Mol Cell *35*, 868-880.

Fan, D.P., Higa, A., and Levinthal, C. (1964). Messenger Rna Decay and Protection. J Mol Biol *8*, 210-222.

Farh, K.K., Grimson, A., Jan, C., Lewis, B.P., Johnston, W.K., Lim, L.P., Burge, C.B., and Bartel, D.P. (2005). The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science *310*, 1817-1821.

Fernandez, J., Yaman, I., Merrick, W.C., Koromilas, A., Wek, R.C., Sood, R., Hensold, J., and Hatzoglou, M. (2002). Regulation of internal ribosome entry site-mediated translation by eukaryotic initiation factor-2alpha phosphorylation and translation of a small upstream open reading frame. J Biol Chem *277*, 2050-2058.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Forstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. (2005). Normal microRNA maturation and germ-line stem cell

maintenance requires Loquacious, a double-stranded RNA-binding domain protein. PLoS Biol *3*, e236.
Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res *19*, 92-105.
Furuichi, Y., Morgan, M., Muthukrishnan, S., and Shatkin, A.J. (1975). Reovirus messenger RNA contains a methylated, blocked 5'-terminal structure: m-7G(5')ppp(5')G-MpCp. Proc Natl Acad Sci U S A *72*, 362-366.
Gallie, D.R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Dev *5*, 2108-2116.
Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. Nat Rev Mol Cell Biol *8*, 113-126.
Gatfield, D., Le Martelot, G., Vejnar, C.E., Gerlach, D., Schaad, O., Fleury-Olela, F., Ruskeepaa, A.L., Oresic, M., Esau, C.C., Zdobnov, E.M., *et al.* (2009). Integration of microRNA miR-122 in hepatic circadian gene expression. Genes Dev *23*, 1313-1326.
Gerber, A.P., Herschlag, D., and Brown, P.O. (2004). Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biol *2*, E79.
Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z., *et al.* (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.
Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat Rev Genet *10*, 94-108.
Gingras, A.C., Raught, B., and Sonenberg, N. (1999). eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. Annu Rev Biochem *68*, 913-963.
Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75-79.
Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199-202.
Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235-240.
Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell *27*, 91-105.
Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-1197.
Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell *106*, 23-34.
Gros, F., Hiatt, H., Gilbert, W., Kurland, C.G., Risebrough, R.W., and Watson, J.D. (1961). Unstable ribonucleic acid revealed by pulse labelling of Escherichia coli. Nature *190*, 581-585.
Gross, P.R., Malkin, L.I., and Moyer, W.A. (1964). Templates for the First Proteins of Embryonic Development. Proc Natl Acad Sci U S A *51*, 407-414.
Grossi de Sa, M.F., Standart, N., Martins de Sa, C., Akhayat, O., Huesca, M., and Scherrer, K. (1988). The poly(A)-binding protein facilitates in vitro translation of poly(A)-rich mRNA. Eur J Biochem *176*, 521-526.
Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587-1590.
Guo, S., and Kemphues, K.J. (1995). par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell *81*, 611-620.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., *et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129-141.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293-296.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev *18*, 3016-3027.

Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887-901.

Hellen, C.U., and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. Genes Dev *15*, 1593-1612.

Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. PLoS Biol *7*, e1000238.

Hentze, M.W., and Kuhn, L.C. (1996). Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. Proc Natl Acad Sci U S A *93*, 8175-8182.

Hinnebusch, A.G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. Annu Rev Microbiol *59*, 407-450.

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., *et al.* (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69-82.

Humphreys, D.T., Westman, B.J., Martin, D.I., and Preiss, T. (2005). MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. Proc Natl Acad Sci U S A *102*, 16961-16966.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet *12*, 99-110.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

Hutvagner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056-2060.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Isken, O., and Maquat, L.E. (2007). Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. Genes Dev *21*, 1833-1856.

Iwasaki, S., Kawamata, T., and Tomari, Y. (2009). Drosophila argonaute1 and argonaute2 employ distinct mechanisms for translational repression. Mol Cell *34*, 58-67.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol *11*, 113-127.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol *3*, 318-356.

Jacobson, A., and Favreau, M. (1983). Possible involvement of poly(A) in protein synthesis. Nucleic Acids Res *11*, 6353-6368.

Jinek, M., Fabian, M.R., Coyle, S.M., Sonenberg, N., and Doudna, J.A. Structural insights into the human GW182-PABC interaction in microRNA-mediated deadenylation. Nat Struct Mol Biol *17*, 238-240.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol *57*, 19-53.

Kahvejian, A., Roy, G., and Sonenberg, N. (2001). The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. Cold Spring Harb Symp Quant Biol *66*, 293-300.

Kasschau, K.D., and Carrington, J.C. (1998). A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing. Cell *95*, 461-470.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., and Siomi, H. (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature *453*, 793-797.

Kedde, M., Strasser, M.J., Boldajipour, B., Oude Vrielink, J.A., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P.M., van Duijse, J., Orom, U.A., *et al.* (2007). RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell *131*, 1273-1286.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev *15*, 2654-2659.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133-141.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.

Kiledjian, M., Wang, X., and Liebhaber, S.A. (1995). Identification of two KH domain proteins in the alpha-globin mRNP stability complex. EMBO J *14*, 4357-4364.

Kim, J., Krichevsky, A., Grad, Y., Hayes, G.D., Kosik, K.S., Church, G.M., and Ruvkun, G. (2004). Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. Proc Natl Acad Sci U S A *101*, 360-365.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126-139.

Kozak, M. (1978). How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell *15*, 1109-1123.

Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. J Biol Chem *266*, 19867-19870.

Kozak, M. (2001). Constraints on reinitiation of translation in mammals. Nucleic Acids Res *29*, 5226-5232.

Krol, J., Busskamp, V., Markiewicz, I., Stadler, M.B., Ribi, S., Richter, J., Duebel, J., Bicker, S., Fehling, H.J., Schubeler, D., *et al.* (2010). Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. Cell *141*, 618-631.

Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M., and Stoffel, M. (2005). Silencing of microRNAs in vivo with 'antagomirs'. Nature *438*, 685-689.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853-858.

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. Curr Biol *14*, 2162-2167.

LaRiviere, F.J., Cole, S.E., Ferullo, D.J., and Moore, M.J. (2006). A late-acting quality control process for mature eukaryotic rRNAs. Mol Cell *24*, 619-626.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858-862.

Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006). Characterization of the piRNA complex from rat testes. Science *313*, 363-367.

Lazzaretti, D., Tournier, I., and Izaurralde, E. (2009). The C-terminal domains of human TNRC6A, TNRC6B, and TNRC6C silence bound transcripts independently of Argonaute proteins. RNA *15*, 1059-1066.

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. Science *294*, 862-864.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

Lee, S.Y., Mendecki, J., and Brawerman, G. (1971). A polynucleotide segment rich in adenylic acid in the rapidly-labeled polyribosomal RNA component of mouse sarcoma 180 ascites cells. Proc Natl Acad Sci U S A *68*, 1331-1335.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., *et al.* (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419.

Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J *21*, 4663-4670.

Leung, A.K., Calabrese, J.M., and Sharp, P.A. (2006). Quantitative analysis of Argonaute protein reveals microRNA-dependent localization to stress granules. Proc Natl Acad Sci U S A *103*, 18125-18130.

Leung, A.K., and Sharp, P.A. (2010). MicroRNA functions in stress responses. Mol Cell *40*, 205-215.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

Li, X., and Carthew, R.W. (2005). A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the Drosophila eye. Cell *123*, 1267-1277.

Li, X., Cassidy, J.J., Reinke, C.A., Fischboeck, S., and Carthew, R.W. (2009). A microRNA imparts robustness against environmental fluctuation during development. Cell *137*, 273-282.

Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769-773.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., and Parker, R. (2005). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. Nat Cell Biol *7*, 719-723.

Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002). Endogenous and silencing-associated small RNAs in plants. Plant Cell *14*, 1605-1619.

Lu, P.D., Harding, H.P., and Ron, D. (2004). Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. J Cell Biol *167*, 27-33.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95-98.

Lytle, J.R., Yario, T.A., and Steitz, J.A. (2007). Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. Proc Natl Acad Sci U S A *104*, 9667-9672.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science *311*, 195-198.

Mallory, A.C., Reinhart, B.J., Jones-Rhoades, M.W., Tang, G., Zamore, P.D., Barton, M.K., and Bartel, D.P. (2004). MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region. EMBO J *23*, 3356-3364.

Mangus, D.A., Evans, M.C., and Jacobson, A. (2003). Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. Genome Biol *4*, 223.

Maroney, P.A., Yu, Y., Fisher, J., and Nilsen, T.W. (2006). Evidence that microRNAs are associated with translating messenger RNAs in human cells. Nat Struct Mol Biol *13*, 1102-1107.
Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563-574.
Mathews, M., Sonenberg, N., and Hershey, J.W.B. (2007). Translational control in biology and medicine, 3rd edn (Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press).
Mathonnet, G., Fabian, M.R., Svitkin, Y.V., Parsyan, A., Huck, L., Murata, T., Biffo, S., Merrick, W.C., Darzynkiewicz, E., Pillai, R.S., *et al.* (2007). MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. Science *317*, 1764-1767.
Melamed, D., and Arava, Y. (2007). Genome-wide analysis of mRNA polysomal profiles with spotted DNA microarrays. Methods Enzymol *431*, 177-201.
Meyuhas, O. (2000). Synthesis of the translational apparatus is regulated at the translational level. Eur J Biochem *267*, 6321-6330.
Morris, D.R., and Geballe, A.P. (2000). Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol *20*, 8635-8642.
Moss, E.G., Lee, R.C., and Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. Cell *88*, 637-646.
Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev *16*, 720-728.
Munroe, D., and Jacobson, A. (1990). mRNA poly(A) tail, a 3' enhancer of translational initiation. Mol Cell Biol *10*, 3441-3455.
Nakamura, A., Sato, K., and Hanyu-Nakamura, K. (2004). Drosophila cup is an eIF4E binding protein that associates with Bruno and regulates oskar mRNA translation in oogenesis. Dev Cell *6*, 69-78.
Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell *2*, 279-289.
Nelson, M.R., Leidal, A.M., and Smibert, C.A. (2004). Drosophila Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression. EMBO J *23*, 150-159.
Nielsen, C.B., Shomron, N., Sandberg, R., Hornstein, E., Kitzman, J., and Burge, C.B. (2007). Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. RNA *13*, 1894-1910.
Nissan, T., and Parker, R. (2008). Computational analysis of miRNA-mediated repression of translation: implications for models of translation initiation inhibition. RNA *14*, 1480-1491.
Nottrott, S., Simard, M.J., and Richter, J.D. (2006). Human let-7a miRNA blocks protein production on actively translating polyribosomes. Nat Struct Mol Biol *13*, 1108-1114.
Nykanen, A., Haley, B., and Zamore, P.D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell *107*, 309-321.
Okamura, K., Balla, S., Martin, R., Liu, N., and Lai, E.C. (2008a). Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. Nat Struct Mol Biol *15*, 998.
Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008b). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature *453*, 803-806.
Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell *130*, 89-100.
Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Dev Biol *216*, 671-680.
Ostareck, D.H., Ostareck-Lederer, A., Shatsky, I.N., and Hentze, M.W. (2001). Lipoxygenase mRNA silencing in erythroid differentiation: The 3'UTR regulatory complex controls 60S ribosomal subunit joining. Cell *104*, 281-290.

Ostareck, D.H., Ostareck-Lederer, A., Wilm, M., Thiele, B.J., Mann, M., and Hentze, M.W. (1997). mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. Cell *89*, 597-606.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P., *et al.* (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature *408*, 86-89.

Pelletier, J., and Sonenberg, N. (1985). Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. Cell *40*, 515-526.

Pestova, T.V., and Kolupaeva, V.G. (2002). The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. Genes Dev *16*, 2906-2922.

Peters, L., and Meister, G. (2007). Argonaute proteins: mediators of RNA silencing. Mol Cell *26*, 611-623.

Petersen, C.P., Bordeleau, M.E., Pelletier, J., and Sharp, P.A. (2006). Short RNAs repress translation after initiation in mammalian cells. Mol Cell *21*, 533-542.

Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. Science *309*, 1573-1576.

Piron, M., Vende, P., Cohen, J., and Poncet, D. (1998). Rotavirus RNA-binding protein NSP3 interacts with eIF4GI and evicts the poly(A) binding protein from eIF4F. EMBO J *17*, 5811-5821.

Rehwinkel, J., Behm-Ansmant, I., Gatfield, D., and Izaurralde, E. (2005). A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. RNA *11*, 1640-1647.

Reinhart, B.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science *297*, 1831.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-906.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev *16*, 1616-1626.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. Cell *110*, 513-520.

Richter, J.D. (2007). CPEB: a life in translation. Trends Biochem Sci *32*, 279-285.

Romano, N., and Macino, G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. Mol Microbiol *6*, 3343-3353.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell *127*, 1193-1207.

Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83-86.

Sachs, A.B., Sarnow, P., and Hentze, M.W. (1997). Starting at the beginning, middle, and end: translation initiation in eukaryotes. Cell *89*, 831-838.

Sachs, M.S., Wang, Z., Gaba, A., Fang, P., Belk, J., Ganesan, R., Amrani, N., and Jacobson, A. (2002). Toeprint analysis of the positioning of translation apparatus components at initiation and termination codons of fungal mRNAs. Methods *26*, 105-114.

Schwartz, D.C., and Parker, R. (1999). Mutations in translation initiation factors lead to increased rates of deadenylation and decapping of mRNAs in Saccharomyces cerevisiae. Mol Cell Biol *19*, 5247-5256.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199-208.

Seggerson, K., Tang, L., and Moss, E.G. (2002). Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation. Dev Biol *243*, 215-225.

Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58-63.
Sen, G.L., and Blau, H.M. (2005). Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. Nat Cell Biol *7*, 633-636.
Shine, J., and Dalgarno, L. (1974). The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc Natl Acad Sci U S A *71*, 1342-1346.
Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol *12*, 246-258.
Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science *305*, 1434-1437.
Stark, A., Brennecke, J., Bushati, N., Russell, R.B., and Cohen, S.M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell *123*, 1133-1146.
Stein, P., Zeng, F., Pan, H., and Schultz, R.M. (2005). Absence of non-specific effects of RNA interference triggered by long double-stranded RNA in mouse oocytes. Dev Biol *286*, 464-471.
Steitz, J.A. (1969). Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. Nature *224*, 957-964.
Steitz, J.A., and Jakes, K. (1975). How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in Escherichia coli. Proc Natl Acad Sci U S A *72*, 4734-4738.
Sulston, J.E., and Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Dev Biol *56*, 110-156.
Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. Science *332*, 472-474.
Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123-132.
Takanami, M., Yan, Y., and Jukes, T.H. (1965). Studies on the site of ribosomal binding of f2 bacteriophage RNA. J Mol Biol *12*, 761-773.
Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., *et al.* (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534-538.
Terenin, I.M., Dmitriev, S.E., Andreev, D.E., and Shatsky, I.N. (2008). Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. Nat Struct Mol Biol *15*, 836-841.
Thermann, R., and Hentze, M.W. (2007). Drosophila miR2 induces pseudo-polysomes and inhibits translation initiation. Nature *447*, 875-878.
Todd, D.J., Lee, A.H., and Glimcher, L.H. (2008). The endoplasmic reticulum stress response in immunity and autoimmunity. Nat Rev Immunol *8*, 663-674.
van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N., and Stuitje, A.R. (1990). Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell *2*, 291-299.
van Rooij, E., Sutherland, L.B., Qi, X., Richardson, J.A., Hill, J., and Olson, E.N. (2007). Control of stress-dependent cardiac growth and gene expression by a microRNA. Science *316*, 575-579.
Vardy, L., and Orr-Weaver, T.L. (2007). Regulating translation of maternal messages: multiple repression mechanisms. Trends Cell Biol *17*, 547-554.
Vattem, K.M., and Wek, R.C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. Proc Natl Acad Sci U S A *101*, 11269-11274.
Vende, P., Piron, M., Castagne, N., and Poncet, D. (2000). Efficient translation of rotavirus mRNA requires simultaneous interaction of NSP3 with the eukaryotic translation initiation factor eIF4G and the mRNA 3' end. J Virol *74*, 7064-7071.

Wakiyama, M., Takimoto, K., Ohara, O., and Yokoyama, S. (2007). Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. Genes Dev *21*, 1857-1862.

Wang, B., Yanez, A., and Novina, C.D. (2008). MicroRNA-repressed mRNAs contain 40S but not 60S components. Proc Natl Acad Sci U S A *105*, 5343-5348.

Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. Trends Biochem Sci *24*, 437-440.

Warner, J.R., Knopf, P.M., and Rich, A. (1963). A multiple ribosomal structure in protein synthesis. Proc Natl Acad Sci U S A *49*, 122-129.

Wassenegger, M., Heimes, S., Riedel, L., and Sanger, H.L. (1994). RNA-directed de novo methylation of genomic sequences in plants. Cell *76*, 567-576.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., *et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539-543.

Wei, C.M., Gershowitz, A., and Moss, B. (1975). Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. Cell *4*, 379-386.

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell *2*, 135-140.

Wightman, B., Burglin, T.R., Gatto, J., Arasu, P., and Ruvkun, G. (1991). Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development. Genes Dev *5*, 1813-1824.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-862.

Wolin, S.L., and Walter, P. (1988). Ribosome pausing and stacking during translation of a eukaryotic mRNA. EMBO J *7*, 3559-3569.

Wu, L., Fan, J., and Belasco, J.G. (2006). MicroRNAs direct rapid deadenylation of mRNA. Proc Natl Acad Sci U S A *103*, 4034-4039.

Xiao, C., Calado, D.P., Galler, G., Thai, T.H., Patterson, H.C., Wang, J., Rajewsky, N., Bender, T.P., and Rajewsky, K. (2007). MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. Cell *131*, 146-159.

Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A. (2003). The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. Curr Biol *13*, 790-795.

Yamashita, A., Chang, T.C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C.Y., and Shyu, A.B. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat Struct Mol Biol *12*, 1054-1063.

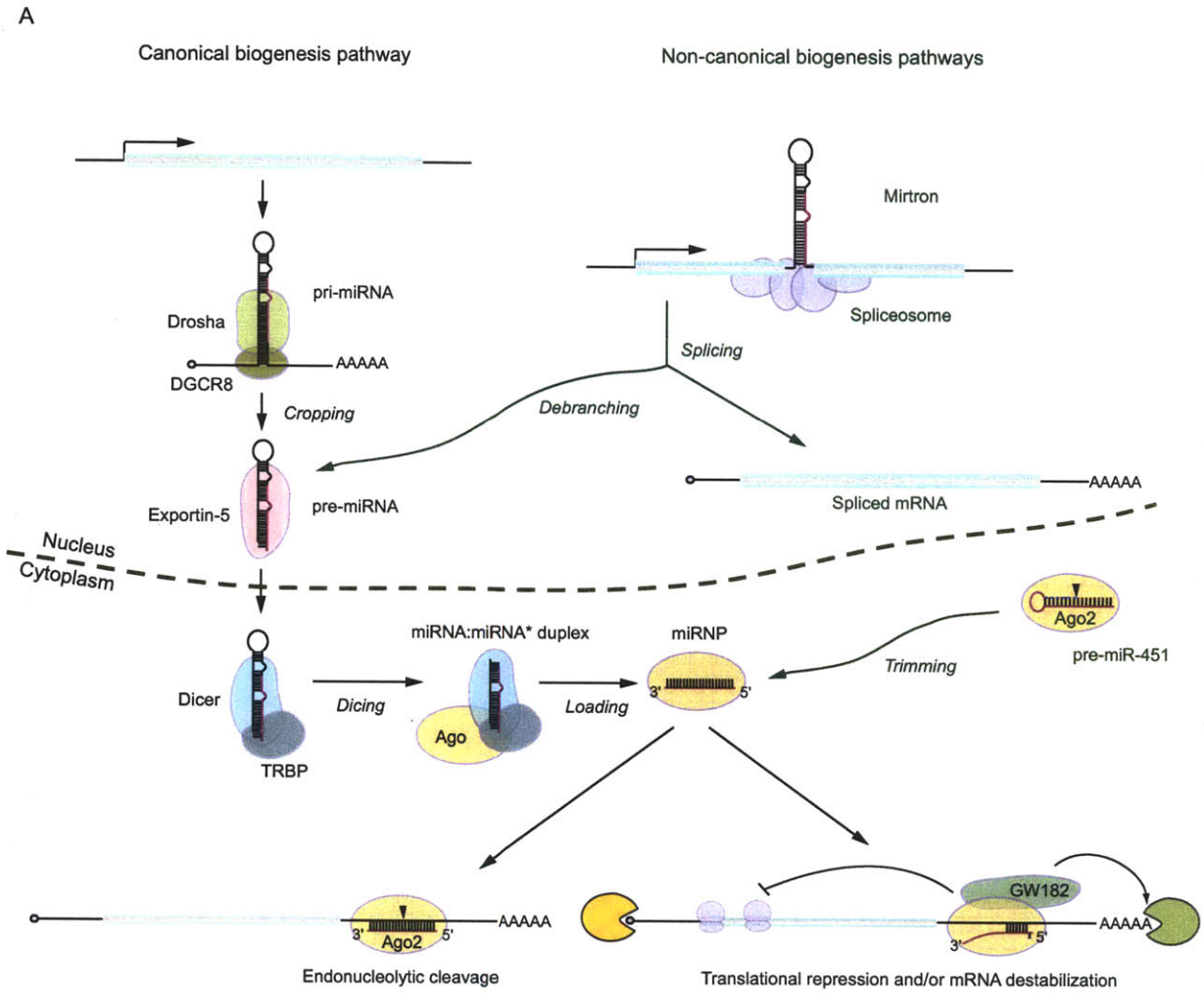
Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science *304*, 594-596.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25-33.

Zekri, L., Huntzinger, E., Heimstadt, S., and Izaurralde, E. (2009). The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. Mol Cell Biol *29*, 6220-6231.

Zeng, Y., Yi, R., and Cullen, B.R. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. Proc Natl Acad Sci U S A *100*, 9779-9784.

Figure 1


A
Canonical biogenesis pathway
Non-canonical biogenesis pathways
Mirtron
Spliceosome
Splicing
pri-miRNA
Drosha
AAAAA
DGCR8
Cropping
Debranching
Exportin-5
pre-miRNA
Spliced mRNA
AAAAA
Nucleus
Cytoplasm
Ago2
pre-miR-451
Trimming
miRNA:miRNA* duplex
miRNP
Dicer
Dicing
Loading
Ago
TRBP
3'
5'
GW182
AAAAA
Endonucleolytic cleavage
Translational repression and/or mRNA destabilization

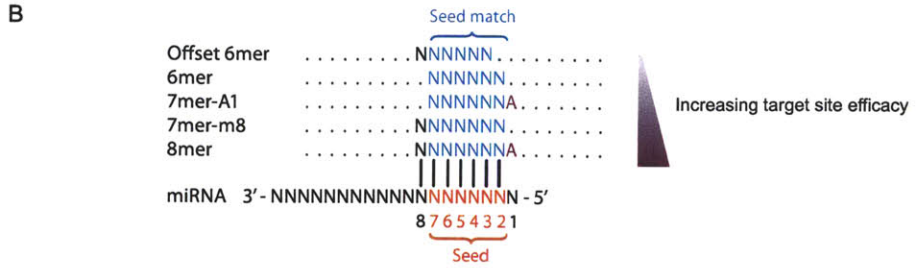
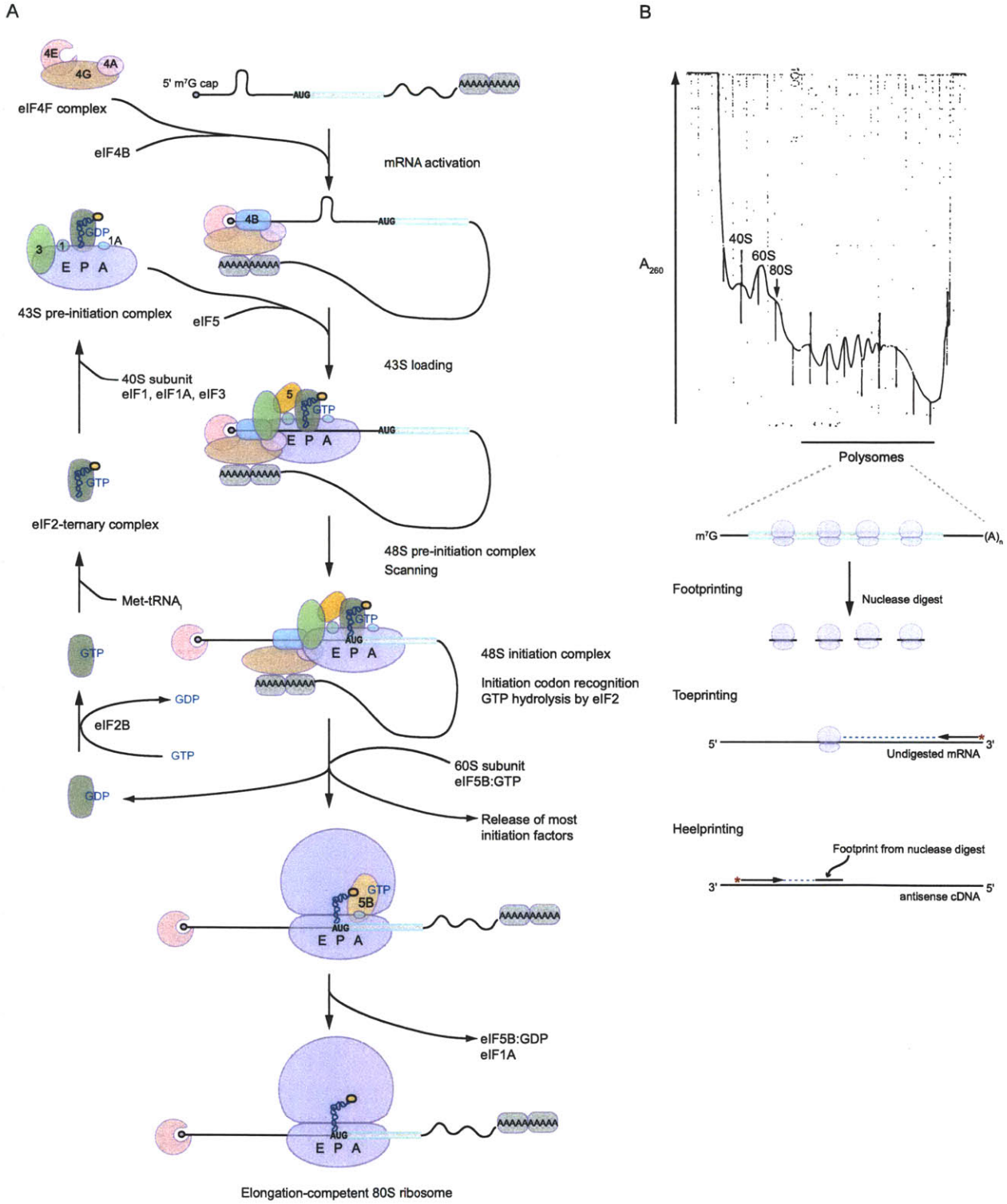

B
Seed match
Offset 6mer
6mer
7mer-A1
7mer-m8
8mer
NNNNNN
NNNNNNA
miRNA 3′- NNNNNNNNNNNNNNNNNNN - 5′
8 7 6 5 4 3 2 1
Seed
Increasing target site efficacy

Figure 2


A
4E
4G
4A
eIF4F complex
5' m7G cap
AUG
eIF4B
mRNA activation
4B
GDP
1A
3
1
E P A
43S pre-initiation complex
eIF5
43S loading
40S subunit
eIF1, eIF1A, eIF3
5
GTP
eIF2-ternary complex
48S pre-initiation complex
Scanning
Met-tRNAi
48S initiation complex
Initiation codon recognition
GTP hydrolysis by eIF2
eIF2B
60S subunit
eIF5B:GTP
Release of most
initiation factors
5B
eIF5B:GDP
eIF1A
Elongation-competent 80S ribosome
B
A260
40S
60S
80S
Polysomes
m7G
(A)n
Footprinting
Nuclease digest
Toeprinting
5'
3'
Undigested mRNA
Heelprinting
Footprint from nuclease digest
antisense cDNA

Chapter 2

Mammalian microRNAs predominantly act to decrease target mRNA levels

Huili Guo[1,2,3], Nicholas T. Ingolia[1,4,5], Jonathan S. Weissman[1,4,5], and David P. Bartel[1,2,3]

[1]Howard Hughes Medical Institute
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[4]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94158, USA
[5]California Institute for Quantitative Biosciences, San Francisco, CA 94158, USA.

H.G. performed the experiments and analyzed the data, with input from the other authors. H.G., J.S.W., and D.P.B. contributed to the design of the study, and all authors contributed to preparation of the manuscript.

Published as:
Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466, 835-840.

MicroRNAs (miRNAs) are endogenous ~22-nucleotide RNAs that mediate important gene-regulatory events by pairing to the mRNAs of protein-coding genes to direct their repression. Repression of these regulatory targets leads to decreased translational efficiency and/or decreased mRNA levels, but the relative contributions of these two outcomes have been largely unknown, particularly for endogenous targets expressed at low-to-moderate levels. Here, we use ribosome profiling to measure the overall effects on protein production and compare these to simultaneously measured effects on mRNA levels. For both ectopic and endogenous miRNA regulatory interactions, lowered mRNA levels account for most (≥84%) of the decreased protein production. These results show that changes in mRNA levels closely reflect the impact of miRNAs on gene expression and indicate that destabilization of target mRNAs is the predominant reason for reduced protein output.

Each highly conserved mammalian miRNA typically targets mRNAs of hundreds of distinct genes, such that as a class these small regulatory RNAs dampen the expression of most protein-coding genes to optimize their expression patterns (Bartel, 2009; Friedman et al., 2009). When pairing to a target is extensive, a miRNA can direct destruction of the targeted mRNA through Argonaute-catalyzed mRNA cleavage (Hutvagner and Zamore, 2002; Liu et al., 2004). This mode of repression dominates in plants (Jones-Rhoades et al., 2006), but in animals all but a few targets lack the extensive pairing required for cleavage (Bartel, 2009).

The molecular consequences of the repression mode that dominates in animals are less clear. Initially miRNAs were thought to repress protein output with little or no influence on mRNA levels (Olsen and Ambros, 1999; Wightman et al., 1993). Then mRNA-array experiments revealed that miRNAs decrease the levels of many targeted mRNAs (Giraldez et al., 2006; Krutzfeldt et al., 2005; Lim et al., 2005; Rehwinkel et al., 2006). A revisit of the initially identified targets of *Caenorhabditis elegans* miRNAs showed that these transcripts also decrease in the presence of their cognate miRNAs (Bagga et al., 2005). The mRNA decreases are associated with poly(A)-tail shortening, leading to a model in which

miRNAs cause mRNA deadenylation, which promotes de-capping and more rapid degradation through standard mRNA-turnover processes (Behm-Ansmant et al., 2006; Eulalio et al., 2009; Giraldez et al., 2006; Wu et al., 2006). The magnitude of this destabilization, however, is usually quite modest, which has bolstered the lingering notion that with some exceptions [e.g., *Drosophila* miR-12 regulation of CG10011 (Behm-Ansmant et al., 2006)] most repression occurs through translational repression, and that monitoring mRNA destabilization might miss many targets that are downregulated without detectable mRNA changes. Challenging this view are results of high-throughput analyses comparing protein and mRNA changes after introducing or deleting individual miRNAs (Baek et al., 2008; Selbach et al., 2008). An interpretation of these results is that the modest mRNA destabilization imparted by each miRNA:target interaction represents most of the miRNA-mediated repression (Baek et al., 2008). We call this the "mRNA-destabilization" scenario and contrast it to the original "translational-repression" scenario, which posited decreased translation with relatively little mRNA change.

In the mRNA-destabilization scenario differences between protein and mRNA changes are mostly attributed to either measurement noise or complications arising from pre-steady-state comparisons of mRNA-array data, which measure differences at one moment in time, and proteomic data, which measure differences integrated over an extended period of protein synthesis. If either mRNA levels or miRNA activities change over the period of protein synthesis (or the period of metabolic labeling), correspondence between mRNA destabilization and protein decreases could become distorted. Another complication of proteomic datasets is that they preferentially examine more highly expressed proteins, whose repression might differ from more modestly expressed proteins. A recent study used mRNA arrays to monitor effects on both mRNA levels and mRNA ribosome density and occupancy, thereby providing a more sensitive analysis of changes in mRNA utilization and bypassing the need to compare protein and mRNA (Hendrickson et al., 2009). This array study supports the mRNA-destabilization scenario but examines the response to an ectopically introduced miRNA, leaving open the question of whether endogenous miRNA:target interactions might impart additional translational repression.

Ribosome profiling, a method that determines the positions of ribosomes on cellular mRNAs with sub-codon resolution (Ingolia et al., 2009), is based on deep sequencing of ribosome-protected mRNA fragments (RPFs) and thereby provides quantitative data on thousands of genes not detected by general proteomics methods. Moreover, ribosome profiling reports on the status of the cell at a particular time point, and thus generates results more directly comparable to mRNA-profiling results than does proteomics. We extended this method to human and mouse cells, thereby enabling a fresh look at the molecular consequences of miRNA repression.

Ribosome profiling in mammalian cells

Ribosome profiling generates short sequence tags that each mark the mRNA coordinates of one bound ribosome (Ingolia et al., 2009). The outline of our protocol for mammalian cells paralleled that used for yeast (Fig. 1a). Cells were treated with cycloheximide to arrest translating ribosomes. Extracts from these cells were then treated with RNase I to degrade regions of mRNAs not protected by ribosomes. The resulting 80S monosomes, many of which contained a ~30-nucleotide RPF, were purified on sucrose gradients and then treated to release the RPFs, which were processed for Illumina high-throughput sequencing.

We started with HeLa cells, performing ribosome profiling on miRNA- and mock-transfected cells. In parallel, poly(A)-selected mRNA from each sample was randomly fragmented, and the resulting mRNA fragments were processed for sequencing (mRNA-Seq) using the same protocol as that used for the RPFs. Sequencing generated 11–18 million raw reads per sample, of which 4–8 million were used for subsequent analyses because they each mapped to a single location in a database of annotated pre-mRNAs and mRNA splice junctions (Supplementary Table 1).

Combining RPFs from HeLa-expressed mRNAs into one composite mRNA showed that ribosome profiling captured fundamental features of translation (Fig. 1b, c and Supplementary Fig. 1c). Although a few RPFs mapped to annotated 5'-untranslated regions (5'UTRs), which indicated the presence of ribosomes at upstream open reading frames (ORFs)(Ingolia et al., 2009), the vast majority

mapped to annotated ORFs. RPF density was highest at the start and stop codons, reflecting known pauses at these positions (Sachs et al., 2002). mRNA-Seq tags, in contrast, mapped uniformly across the length of the mRNA, as expected for randomly fragmented mRNA.

The most striking feature in the composite-mRNA analysis was the 3-nucleotide periodicity of the RPFs. In sharp contrast to the 5' termini of the mRNA-Seq tags, which mapped to all three codon nucleotides equally, the RPF 5' termini mostly mapped to the first nucleotide of the codon (Fig. 1d). This pattern, analogous to that observed in yeast (Ingolia et al., 2009), is attributable to the RPFs capturing the movement of ribosomes along mRNAs—three nucleotides at a time. The protocol applied to mouse neutrophils generated ~30-nucleotide RPFs with the same pattern (Supplementary Fig. 1d, e). Thus, ribosome profiling mapped, at sub-codon resolution, the positions of translating ribosomes in human and mouse cells.

Similar repression regardless of target expression level

General features of translation and translational efficiency in mammalian cells will be presented elsewhere. Here, we focus on miRNA-dependent changes in protein production. Our HeLa-cell experiments examined the impact of introducing miR-1 or miR-155, both of which are not normally expressed in HeLa cells, and our mouse-neutrophil experiments examined the impact of knocking out *mir-223*, which encodes a miRNA highly and preferentially expressed in neutrophils (Johnnidis et al., 2008). These cell types and miRNAs were chosen because proteomics experiments using either the SILAC (stable isotope labeling with amino acids in cell culture) or pSILAC (a pulsed-labeled version of SILAC) methods had already reported the impact of each of these miRNAs on the output of thousands of proteins (Baek et al., 2008; Selbach et al., 2008).

Pairing to the miRNA seed (nucleotides 2–7) is important for target recognition, and several types of seed-matched sites, ranging in length from 6 to 8 nucleotides, mediate repression (Bartel, 2009). Ribosome-profiling and mRNA-Seq results showed the expected correlation between site length and site efficacy (Bartel, 2009) (Supplementary Fig. 2). Because the response of mRNAs with single 6-nucleotide

sites was marginal and observed only in the miR-1 experiment, subsequent analyses focused on mRNAs with at least one canonical 7–8-nucleotide site.

In the miR-155 experiment, mRNAs from 5,103 distinct genes passed our read threshold for single-gene quantification (≥100 RPFs and ≥100 mRNA-Seq tags in the mock-transfection control). Genes with at least one 3'UTR site tended to be repressed following addition of miR-155, yielding fewer mRNA-Seq tags and fewer RPFs in the presence of the miRNA [Fig. 2a; $P < 10^{-48}$ and 10^{-37}, respectively, one-tailed Kolmogorov–Smirnov (K–S) test, comparing to genes with no site in the entire message]. Proteins from 2,597 of the 5,103 genes were quantified in the analogous pSILAC experiment (Selbach et al., 2008). The mRNA and RPF changes for the pSILAC-detected subset were no less pronounced than those of the larger set of analyzed genes (Fig. 2a; $P = 0.70$ and 0.62 for mRNA and RPF data, respectively, K–S test), which implied that the response of mRNAs of proteins detected by high-throughput quantitative proteomics accurately represented the response of all mRNAs. Analogous results were obtained in the miR-1 and miR-223 experiments (Fig. 2b, c; $P < 10^{-10}$ for each comparison to genes with no site, and $P > 0.56$ for each comparison to the proteomics-detected subset). Furthermore, analyses of genes binned by expression level, which enabled inclusion of data from 11,000 distinct genes that ranged broadly in expression (more than 1,000-fold difference between the first and last bins), confirmed that miRNAs do not repress their lowly expressed targets more potently than they do their more highly expressed targets (Supplementary Fig. 3).

As these results indicated that restricting analyses to mRNAs with higher expression, by requiring either a minimal read count or a proteomics-detected protein, did not somehow distort the picture of miRNA targeting and repression, we focused on the mRNAs with at least one 3'UTR site and for which the proteomics detected a substantial change at the protein level. These sets of mRNAs were called "proteomics-supported targets" because they were expected to be highly enriched in direct targets of the miRNAs. Indeed, they responded more robustly to the introduction or ablation of cognate miRNAs (Fig. 2a–c; $P < 10^{-5}$ for each comparison to proteomics-detected genes with sites). Because some 7–8-nucleotide seed-matched sites do not confer repression by the corresponding miRNA (Bartel, 2009;

Grimson et al., 2007), the proteomics-supported targets, which excluded most messages with nonfunctional sites, were the most informative for subsequent analyses.

Modest influence on translational efficiency

We next examined whether our results supported the translation-repression scenario, in which translation is repressed without a substantial mRNA decrease. In the characterized examples in which miRNAs direct translation inhibition, repression is reported to occur through either reduced translation initiation (Chendrimada et al., 2007; Humphreys et al., 2005; Pillai et al., 2005) or increased ribosome drop-off (Petersen et al., 2006). Both of these mechanisms would lead to fewer ribosomes on target mRNAs and thus fewer RPFs from these mRNAs after accounting for changes in mRNA levels. To detect this effect, we accounted for changes in mRNA levels by incorporating the mRNA-Seq results. For example, for each quantified gene in the miR-155 experiment, we divided the change in RPFs by the change in mRNA-Seq tags (i.e., we subtracted the $\log_2$-fold changes). This calculation removed the component of the RPF change attributable to miRNA-dependent changes in poly(A) mRNA, leaving the residual change as the component attributable to a change in ribosome density, which we interpret as a change in "translational efficiency (Ingolia et al., 2009)".

We observed a statistically significant decrease in translational efficiency for messages with miR-155 sites compared to those without, indicating that miRNA targeting leads to fewer ribosomes on target mRNAs that have not yet lost their poly(A)-tail and become destabilized (Fig. 2d, $P = 0.003$, K–S test). This decrease, however, was very modest. Even these proteomics-supported targets underwent only a 7% decrease in translational efficiency (–0.11 $\log_2$-fold change, Fig. 2d, inset), compared to a 33% decrease in polyadenylated mRNA (–0.59 $\log_2$-fold change, Fig. 2a). Analogous results were obtained for the miR-1 and miR-223 experiments (Fig. 2e, f; $P = 0.001$, $P = 0.05$, respectively). Thus, for both ectopic and endogenous regulatory interactions, only a small fraction of repression observed by ribosome profiling (11–16%) was attributable to reduced translational efficiency. At least 84% of the repression

was attributable instead to decreased mRNA levels, a percentage somewhat greater than the ~75% reported from array analyses of ectopic interactions (Hendrickson et al., 2009).

Analyses described thus far focused on messages with at least one 3'UTR site to the cognate miRNA, without considering whether or not the site was conserved in orthologous UTRs of other animals. When we focused on evolutionarily conserved sites (Friedman et al., 2009), the results were similar but noisier because the conserved sites, although more efficacious, were 3–13-fold less abundant (Supplementary Fig. 4). When changing the focus to messages with sites only in the ORFs, the results were also similar but again noisier because sites in the open reading frames are less efficacious (Baek et al., 2008; Grimson et al., 2007; Selbach et al., 2008), which led to ~70% fewer genes classified as proteomics-supported targets (Supplementary Fig. 5).

mRNA reduction consistently mirrored RPF reduction

Analyses of fold-change distributions (Fig. 2) supported the mRNA-destabilization scenario for most targets, but still allowed for the possibility that the translational-repression scenario might apply to a small subset of targets. To search for evidence for a set of unusual targets undergoing translational repression without substantial mRNA destabilization, we compared the mRNA and ribosome-profiling changes for the 5,103 quantifiable genes from the miR-155 experiment. Correlation between the two types of responses was strong for the messages with miR-155 sites, and particularly for those that were proteomics-supported targets (Fig. 3a, R^2 = 0.49 and 0.63, respectively). A strong correlation was also observed for genes considered only after relaxing the expression cutoffs (Supplementary Fig. 6a). Any scatter that might have indicated that a few genes undergo translational repression without substantial mRNA destabilization strongly resembled the scatter observed in parallel analysis of genes without sites (Fig. 3b). The same was observed for the miR-1 experiment, but in this case the correlations were even stronger (R^2 = 0.72 and 0.80, respectively), presumably because the increased response to the miRNA led to a correspondingly reduced contribution of experimental noise (Fig. 3c, d; Supplementary Fig. 6b). The same was also observed for the miR-223 experiment, with weaker correlations (R^2 = 0.26 and 0.40,

respectively) attributable to the reduced response to the miRNA and a correspondingly increased contribution of experimental noise (Fig. 3e, f). Supporting this interpretation, systematically increasing expression cutoffs, which retained data with progressively lower noise from stochastic counting fluctuations, progressively increased the correlation between RPF and mRNA-Seq changes (Supplementary Fig. 6c). We also examined messages with multiple sites to the cognate miRNA and found that they behaved no differently with regard to the relationship between mRNA-Seq and RPF changes (Supplementary Fig. 7). In summary, we found no evidence that countered the conclusion that miRNAs predominantly act to reduce mRNA levels of nearly all, if not all, targets.

Uniform changes along the ORF length

If miRNA targeting causes ribosomes to drop off the message after translating a substantial fraction of the ORF, then the RPF changes summed over the length of the ORF might underestimate the reduced production of full-length protein. Therefore, we re-examined the ribosome profiling data, which determines the location of ribosomes along the length of the mRNAs, thereby providing transcriptome-wide information that could detect ribosome drop-off. For highly expressed genes targeted in their 3'UTRs (e.g, *TAGLN2* in the miR-1 experiment; Supplementary Fig. 8a), downregulation at the mRNA and ribosome levels was observed along the length of the ORF. In order to extend this analysis to genes with more moderate expression, we examined composite ORFs representing proteomics-supported targets and compared these to composite ORFs representing genes without sites. When miR-155 targets were compared to genes without sites, fewer mRNA-Seq tags were observed across the length of the composite ORF (Fig. 4a). RPFs tended to be further reduced ($P = 0.007$, one-tailed Mann–Whitney test), but without a systematic change in the magnitude of this additional reduction across the length of the ORF [$P = 0.95$, two-tailed Analysis of Covariance (ANCOVA) test]. Because ribosome drop-off would decrease the ribosome occupancy less at the beginning of the ORF than at the end, whereas inhibiting translation initiation would not, the observed uniform reduction supported mechanisms in which initiation was inhibited. Analogous results were observed in the miR-1 experiment (Fig. 4b; $P = 0.002$, for further

reduction in RPFs; $P = 0.85$ for systematic change across the ORF). Evidence for drop-off was also not observed in the miR-223 experiment, although a change in translational efficiency was difficult to detect in this analysis, presumably because the miRNA-mediated changes were lower in magnitude (Fig. 4c). The same conclusions were drawn from analyses in which we first normalized for ORF length (Supplementary Fig. 9).

Implications for the mechanism of repression

For both ectopic and endogenous miRNA targeting interactions, the molecular consequences of miRNA regulation were most consistent with the mRNA-destabilization scenario. Although acquiring similar data on cell types beyond the two examined here will be important, we have no reason to doubt that our conclusion will apply broadly to the vast majority of miRNA targeting interactions. If indeed general, this conclusion will be welcome news to biologists wanting to measure the ultimate impact of miRNAs on their direct regulatory targets. Because the quantitative effects on translating ribosomes so closely mirrored the decreases in polyadenylated mRNA, the impact on protein production can be closely approximated using mRNA arrays or mRNA-Seq. Our results might also provide insight into the question of why some targets are more responsive to miRNAs than others; in the destabilization scenario, otherwise long-lived messages might undergo comparatively more destabilization than would constitutively short-lived ones.

Translation repression and mRNA destabilization are sometimes coupled (Coller and Parker, 2004), which raises the possibility that the miRNA-mediated mRNA destabilization might be a consequence of translational repression. If so, a greater fraction of the repression might be attributable to decreased translational efficiency if the effects were analyzed sooner after introducing a miRNA. However, the fraction attributable to decreased translational efficiency remained small when repeating the analysis using samples from 12 hours (rather than 32 hours) after introducing miR-155 or miR-1 (Supplementary Fig. 10 and Supplementary Table 2). Although these results at earlier time points cannot rule out rapid destabilization as a consequence of translational repression, our results revealing such small

decreases in translational efficiency for target mRNAs strongly imply that even if destabilization were secondary to translational repression, it would be this destabilization (i.e., the reduced availability of mRNA for subsequent rounds of translation) that would exert the greatest impact on protein production. Moreover, miRNA-mediated mRNA deadenylation, which is the best-characterized mechanism of miRNA-mediated mRNA destabilization, can occur with or without translation of the ORF (Eulalio et al., 2009; Eulalio et al., 2007; Giraldez et al., 2006; Wu et al., 2006), which suggests that the miRNA-mediated destabilization does not result from translational repression and indicates that translational repression could occur after the initial deadenylation signal. Perhaps the miRNA-induced poly(A)-tail interactions that eventually trigger deadenylation also cause the closed circular form of the mRNA to open up, thereby inhibiting translation initiation. This inhibition would occur before deadenylation is complete, as polyadenylated mRNAs seem to be translationally repressed (Fig. 2d–f).

Another consideration is that, as done previously (Baek et al., 2008; Hendrickson et al., 2009; Selbach et al., 2008), we equated mRNA destabilization to the loss of polyadenylated mRNA. Thus, transcripts that have lost their poly(A) tails might still be present but underrepresented in our mRNA-Seq of poly(A)-selected mRNA. In certain cell types, most notably oocytes, such transcripts can be stable and eventually be tailed by a cytoplasmic polyadenylation complex to become translationally competent (Mendez and Richter, 2001). In the typical somatic cell, however, deadenylated transcripts are not translated and are instead rapidly decapped and/or degraded. Thus, our consideration of deadenylated transcripts as operational and functional equivalents of degraded transcripts seems appropriate. One possibility, though, is that mRNAs that were deadenylated while being translated will yield some RPFs from ribosomes that initiated when the poly(A) tails were intact but will not yield mRNA-Seq tags. However, a narrowing of the differences between changes in RPFs and mRNA-Seq tags through this process is expected to have been very small, since the vast majority of RPFs should derive from mRNAs with poly(A) tails.

A way that our results might still be reconciled with the translation-repression scenario would be if ribosome profiling missed the bulk of translation repression because translation was repressed without

reducing the density of ribosomes on the targeted messages, i.e., if reduced initiation was coupled with correspondingly slower elongation. However, direct evidence for slower elongation has not been reported in any miRNA studies, and it seems unlikely that decreases in initiation and elongation rates would so frequently be so closely matched so as to yield such minor differences in apparent translational efficiency for so many messages. Moreover, translational repression without changes in ribosome density would cause the changes measured by proteomics to exceed those measured by ribosome profiling. The same would hold for cotranslational degradation of nascent polypeptides, another proposed mechanism for miRNA-mediated repression (Nottrott et al., 2006; Olsen and Ambros, 1999). Arguing strongly against both of these possibilities, we found that changes measured by proteomics were not greater than those measured by ribosome profiling (Supplementary Fig. 11).

Although the changes we observed in translational efficiency were consistent with slightly reduced translation of the targeted messages, such changes could also occur without any miRNA-mediated translational repression. If some fraction of the polyadenylated mRNA was in a cellular compartment sequestered away from the compartment containing both miRNAs and ribosomes, then preferential destabilization of the mRNA in the miRNA/ribosome compartment would lead to an observed decrease in translational efficiency without a need to invoke translational repression. For example, to the extent that mature mRNAs awaiting transport to the cytoplasm reside in the nucleus where they presumably would not be subject to either miRNA-mediated destabilization or translation, the reduction of mRNA-Seq tags would not match the reduction of RPFs, and the more pronounced RPF reduction would indicate decreased ribosome density even in the absence of translational repression. Heterologous reporter mRNAs, some of which have lent support to the translational-repression scenario, might be particularly prone to nuclear accumulation. With this consideration in mind, the observed miRNA-dependent reductions in translational efficiency might be considered upper limits on the magnitude of translational repression.

Although we cannot determine the precise amount of miRNA-mediated translational repression, we can reliably say that the pervasive and dominant miRNA-mediated translational repression with

persistence of repressed mRNAs, which had been widely anticipated, has not materialized. Instead, the outcome of regulation is predominantly mRNA destabilization, as first suggested by analyses of proteomic data (Baek et al., 2008). We cannot rule out a few interactions for which there is substantial translational repression with little or no mRNA destabilization, but if these exist, they would be rare outliers. For such outliers, miRNAs might be working in concert with other mRNA-binding factors such that the action of the other factors depends on miRNA binding. Such outliers with readily detectable translational repression would be the most attractive subjects of mechanistic studies. The mechanism of translational repression might differ for different messages, depending on the identity of the cooperating factors, perhaps helping to explain the diversity of reported mechanisms by which miRNAs translationally repress their targets (Filipowicz et al., 2008). Understanding these potential elaborations of miRNA-mediated repression would be important, as is a more thorough mechanistic understanding of the predominant reason for reduced protein output, which is mRNA destabilization.

Methods

Transfections and neutrophil culture. HeLa cells were transfected with 100 nM miRNA duplex as described (Selbach et al., 2008). At 12 h and 32 h post-transfection, cycloheximide (100 μg/ml) was added to arrest translation, and after incubating 8 min at 37 °C, cells were harvested in ice-cold PBS supplemented with cycloheximide. For each transfection, cells from six 6-cm dishes were combined and then split into two portions, one for mRNA profiling and the other for ribosome profiling. Haematopoietic progenitors were isolated from two 3-month-old WT male mice and two 3-month-old *mir-223* KO male mice and grown in Iscove's modified Dulbecco's medium (IMDM) containing granulocyte colony-stimulating factor (G-CSF) and stem cell factor (SCF) as described (Baek et al., 2008). On day 6, cycloheximide (100 μg/ml) was added to arrest translation. After incubating 8 min at 37 °C, cells were harvested in ice-cold PBS supplemented with cycloheximide and split into two portions for mRNA profiling and ribosome profiling.

Ribosome footprinting and RPF purification. Cells were pelleted and resuspended in ice-cold lysis buffer (10 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$, 100 mM KCl, 1% Triton X-100, 2 mM DTT, 100 μg/ml cycloheximide, 500 U/ml RNasin, 1 x complete protease inhibitor). The lysis mixture was homogenized six times with a 26-gauge needle at 4 °C and centrifuged at 1,300*g* for 8 min. The supernatant was snap-frozen for later use or processed immediately. RNase I (Ambion, final concentration, 0.5–1.0 U/ul) was added to the cell extract, and the reaction was incubated for 30 min on a shaker at room temperature (~25 °C). Digested extracts were layered onto 11-ml 10–50% linear sucrose gradients that were prepared by horizontal diffusion and centrifuged in an SW-41Ti rotor at 36,000 r.p.m for 2 h. Gradients were fractionated by upward displacement with 60% sucrose on a gradient fractionator (Brandel). Monosome fractions were pooled, concentrated using Ultra-4 centrifugal filters with Ultracel-100 membranes (Amicon) by centrifuging at 1,900*g* for 30 min at 4 °C. Ice-cold release buffer (20 mM HEPES-KOH, 100 mM KCl, 1 mM EDTA, 2 mM DTT, 20 U/ml SUPERase•In, Ambion) was then added to the retentate, and the mixture was incubated on ice for 10 min to release mRNA fragments from ribosomal subunits, after which the mixture was again centrifuged at 1,900*g* for 15 min at 4 °C. The filtrate was then supplemented with SDS to 1% and treated with proteinase K (200 μg/ml) for 30 min at 42 °C. RNA was extracted with acid phenol:chloroform (pH 4.5, Ambion), ethanol-precipitated and resuspended in water. Pilot experiments, using nuclease-protection assays like those performed for yeast samples (Ingolia et al., 2009), showed that the lengths of mammalian RPFs centered at ~30 nucleotides. Therefore, RPFs were gel-purified on a denaturing 10% polyacrylamide-urea gel, excising the region corresponding to 27–33 nucleotides, with the intent of avoiding abundant ribosomal RNA degradation fragments that were 26 and 35 nucleotides in length.

mRNA fragmentation and microarrays. Total RNA was isolated using TRI Reagent (Ambion) and poly(A)$^+$ mRNA was isolated using oligo(dT) DynaBeads (Invitrogen) according to manufacturers' instructions. Alkaline fragmentation buffer (2 mM EDTA, 10 mM Na_2CO_3, 90 mM $NaHCO_3$, pH ≈ 9.3) was added to an equal volume of the purified mRNA and the reaction incubated for 20 min at 95 °C. Ice-

cold stop solution (final 0.3 M NaOAc, pH 5.2, with GlycoBlue co-precipitant, Ambion) was then added, and RNA was ethanol precipitated. RNA fragments from ~25–45 nucleotides were gel-purified on a denaturing 10% polyacrylamide-urea gel. Each sample of total RNA was also analyzed by microarray profiling, using the Affymetrix platform: Human Genome U133 Plus 2.0 Array, or Mouse Genome 430 2.0 Array.

Small-RNA library preparation. Libraries for Illumina sequencing were prepared as described (Grimson et al., 2008) but with the following modifications. Because RPFs and alkaline fragmentation products terminate with a 5'-hydroxyl and a 3'-phosphate, they were 3'-dephosphorylated with polynucleotide kinase (PNK, New England Biolabs) for 6 h at 37 °C in dephosphorylation buffer (100 mM MES-NaOH, pH 5.5, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 300 mM NaCl, 0.5 U/ul enzyme) and desalted (Microspin G-25 column, Amersham) before ligation to the 3' adaptor. Gel-purified 3'-ligation products were then 5'-phosphorylated with PNK, according to manufacturer's instructions, before the 5'-ligation step. Despite steps taken to minimize ribosomal RNA contamination, our ribosome-profiling libraries were initially contaminated by high levels of rRNA (ranging from 60–93%). To enrich for RPFs, DNA from each library was amplified for an additional six cycles and then gel-purified on a 90% formamide, 8% acrylamide denaturing gel. With this additional step, ribosomal RNA contamination was reduced to 40–54%.

Sequence analyses. Illumina sequencing reads were mapped to the reference genome (hg18 for human, mm9 for mouse) with the Bowtie short-read mapping program (Langmead et al., 2009) using the first 25 nucleotides as the 'seed' region. Reads with multiple equivalent hits to the genome were discarded, as were reads that mapped to ribosomal RNA and other annotated noncoding RNAs. To allow for a miscalled residue within the seed region, reads that had failed to map when allowed no seed mismatches were fed into Bowtie again, this time allowing for one seed mismatch. To capture reads uniquely spanning splice junctions, reads that failed to map to the genome were mapped to a set of reference

transcripts, using the same two-stage iterative mapping and again discarding those with multiple equivalent hits. These uniquely transcript-matching reads were combined with the genome-matching reads for subsequent analyses. To compile the set of reference transcripts we started from only curated coding transcripts (entries with NM accession numbers) in the RefSeq database (refFlat files, generated on August 9, 2009, were downloaded from the UCSC Genome Browser, http://genome.ucsc.edu). Of these, transcripts with incomplete coding sequences or those that could be potential substrates of nonsense-mediated decay were filtered out. If a gene had multiple isoforms remaining after this filtering, the longest isoform was picked to represent it. This non-redundant set of mRNAs from unique genes then served as our reference transcript database. Reads of ambiguous origin, such as a read that could derive from either of two different overlapping genes, were discarded. Of the remaining reads, those that could be unambiguously assigned to an exon or intron from a gene represented in our reference transcript database were attributed to that gene. The reference transcript databases for both human and mouse will be available for anonymous download at http://web.wi.mit.edu/bartel/pub/publication.html.

Quantification of gene expression. A modified version of reads per kilobase exon model per million mapped reads (rpkM) was used to quantify gene expression. The original rpkM, developed for RNA-Seq (Mortazavi et al., 2008), was calculated as such: $R = (10^9 \cdot C/N \cdot L)$, where C is the number of mapped reads in a gene's exons, N is the total number of reads mapped (library size), and L is the length of the sum of the exons in nucleotides. To prevent ribosomal RNA contamination in the RPF libraries from skewing our measurements of gene expression, the library size was taken to be the total number of reads mapping to all the exons and introns of our reference transcript database (N'). Because we were interested in comparing mRNA-level and translation-level expression, the length of the open reading frame was taken to be the feature length of each gene (L') and we only included reads mapping to coding exons (C') in our quantification. Hence, rpkM in this study refers to $R' = (10^9 \cdot C'/N' \cdot L')$. Fold changes were calculated by dividing the normalized gene expression value in the experimental condition by the same measure in the control condition. For the cumulative-distribution plots, the median of the distribution of genes without

seed matches (No site) was subtracted from all the fold changes (including those from messages with sites). This normalization caused our reported fold-change distributions of the genes without sites to center on zero. Thresholds for gene quantification, when applied, were applied to the mock transfection data set or the *mir-223* KO data set.

Figure legends

Figure 1. Ribosome profiling in human cells captured features of translation. a, Schematic diagram of ribosome profiling. Sequencing reproducibility and evidence for mapping to the correct mRNA isoforms are illustrated (Supplementary Fig. 1a, b). **b,** RPF density near the ends of ORFs, combining data from all quantified genes. Plotted are RPF 5' termini, as reads per million reads mapping to genes (rpM). Illustrated below the graph are the inferred ribosome positions corresponding to peak RPF densities, at which the start codon was in the P site (*left*) and the stop codon was in the A site (*right*). The offset between the 5' terminus of an RPF and the first nucleotide in the human ribosome A site was typically 15 nucleotides (nt). **c,** Density of RPFs and mRNA-Seq tags near the ends of ORFs in HeLa cells. RPF density is plotted as in panel **b**, except positions are shifted +15 nucleotides to reflect the position of the first nucleotide in the ribosome A site. Composite data are shown for ≥600-nucleotide ORFs that passed our threshold for quantification (≥100 RPFs and ≥100 mRNA-Seq tags). **d,** Fraction of RPFs and mRNA-Seq tags mapping to each of the three codon nucleotides in panel **c**.

Figure 2. MicroRNAs downregulated gene expression mostly through mRNA destabilization, with a small effect on translational efficiency. a, Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right*) after introducing miR-155. Plotted are distributions for the genes with ≥1 miR-155 3'UTR site (blue), the subset of these genes detected in the pSILAC experiment (proteomics-detected, red), the subset of the proteomics-detected genes with proteins responding with $\log_2$-fold change ≤ –0.3 (proteomics-supported, green), and the control genes, which lacked miR-155 sites throughout their mRNAs (no site, black). The number of genes in each category is indicated in parentheses. **b,**

Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right*) after introducing miR-1. Otherwise, as in panel **a**. **c,** Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right*) after deleting *mir-223*. Otherwise, as in panel **a**, with proteomics-supported genes referring to genes with proteins that responded with $\log_2$-fold change ≥ 0.3 in the SILAC experiment. **d,** Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after introducing miR-155. For each gene, the translational efficiency change was calculated by normalizing the RPF change by the mRNA-Seq change. For each distribution, the mean $\log_2$-fold change ($\pm$ standard error) is shown (*inset*). **e,** Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after introducing miR-1. Otherwise, as in panel **d**. **f,** Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after deleting *mir-223*. Otherwise, as in panel **d**.

Figure 3. Ribosome changes from miRNA targeting corresponded to mRNA changes. a, Correspondence between ribosome (RPF) and mRNA (mRNA-Seq) changes after introducing miR-155, plotting data for the 707 quantified genes with at least one miR-155 3'UTR site (blue circles). Proteomics-detected targets and proteomics-supported targets are highlighted (pink diamonds and green crosses, respectively). Expected standard deviations (error bars) were calculated based on the number of reads obtained per gene and assuming random counting statistics. The R^2 derived from Pearson's correlation of all data is indicated. **b,** Correspondence between ribosome and mRNA changes after introducing miR-155, plotting data for 707 genes randomly selected from the 3,186 quantified genes lacking a miR-155 site anywhere in the mRNA. Otherwise, as in panel **a**. **c** and **d,** As in panels **a** and **b**, but plotting results for the miR-1 experiment. **e** and **f,** As in panels **a** and **b**, but plotting results for the miR-223 experiment.

Figure 4. Ribosome and mRNA changes were uniform along the length of the ORFs. a, Ribosome and mRNA changes along the length of ORFs after introducing miR-155. mRNA segments of quantified

genes were binned based on their distance from the first nucleotide of the start codon, with the boundaries of the segments chosen such that each bin contained the same number of nucleotides (Supplementary Fig. 8b). Binning was done separately for mRNAs with no miR-155 site and proteomics-supported miR-155 targets. Fold changes in RPFs and mRNA-Seq tags mapping to each bin were then plotted with respect to the median distance of the central nucleotide of each segment from the first nucleotide of the start codon. Changes in RPFs and mRNA-Seq tags for mRNAs with no site (grey and black, respectively) and for proteomics-supported targets (light and dark green, respectively) are shown. Only bins with read contribution from ≥20 genes are shown (see Supplementary Fig. 8b). The ANCOVA test for systematic change across the ORF length was performed by first calculating the differences between RPF changes and mRNA-Seq changes for each group of genes, fitting lines through these changes in translational efficiency, then testing for a difference between the resulting slopes. **b,** As in panel **a**, but plotting results for the miR-1 experiment. **c,** As in panel **a**, but plotting results for the miR-223 experiment.

Acknowledgements

We thank F. Camargo, C. Jan, J. Kim and C. Petersen for advice and discussions, O. Rissland for comments on the manuscript, and the Whitehead Institute's Genome Technology Core for sequencing and microarray profiling. This work was supported by grants from the NIH (D.P.B. and J.S.W.). H.G. was supported by the Agency for Science, Technology and Research, Singapore. N.T.I. was supported by a Ruth L. Kirschstein National Research Service Award (GM080853). D.P.B and J.S.W. are investigators of the Howard Hughes Medical Institute. Small-RNA sequencing data and array data were deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) under accession number GSE22004.

References

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature *455*, 64-71.

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A.E. (2005). Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. Cell *122*, 553-563.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes Dev *20*, 1885-1898.

Chendrimada, T.P., Finn, K.J., Ji, X., Baillat, D., Gregory, R.I., Liebhaber, S.A., Pasquinelli, A.E., and Shiekhattar, R. (2007). MicroRNA silencing through RISC recruitment of eIF6. Nature *447*, 823-828.

Coller, J., and Parker, R. (2004). Eukaryotic mRNA decapping. Annu Rev Biochem *73*, 861-890.

Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., and Izaurralde, E. (2009). Deadenylation is a widespread effect of miRNA regulation. RNA *15*, 21-32.

Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.F., Doerks, T., Dorner, S., Bork, P., Boutros, M., and Izaurralde, E. (2007). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. Genes Dev *21*, 2558-2570.

Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet *9*, 102-114.

Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res *19*, 92-105.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75-79.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell *27*, 91-105.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-1197.

Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. PLoS Biol *7*, e1000238.

Humphreys, D.T., Westman, B.J., Martin, D.I., and Preiss, T. (2005). MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. Proc Natl Acad Sci U S A *102*, 16961-16966.

Hutvagner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056-2060.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Johnnidis, J.B., Harris, M.H., Wheeler, R.T., Stehling-Sun, S., Lam, M.H., Kirak, O., Brummelkamp, T.R., Fleming, M.D., and Camargo, F.D. (2008). Regulation of

progenitor cell proliferation and granulocyte function by microRNA-223. Nature *451*, 1125-1129.
Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol *57*, 19-53.
Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M., and Stoffel, M. (2005). Silencing of microRNAs in vivo with 'antagomirs'. Nature *438*, 685-689.
Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.
Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769-773.
Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.
Mendez, R., and Richter, J.D. (2001). Translational control by CPEB: a means to the end. Nat Rev Mol Cell Biol *2*, 521-529.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.
Nottrott, S., Simard, M.J., and Richter, J.D. (2006). Human let-7a miRNA blocks protein production on actively translating polyribosomes. Nat Struct Mol Biol *13*, 1108-1114.
Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Dev Biol *216*, 671-680.
Petersen, C.P., Bordeleau, M.E., Pelletier, J., and Sharp, P.A. (2006). Short RNAs repress translation after initiation in mammalian cells. Mol Cell *21*, 533-542.
Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. Science *309*, 1573-1576.
Rehwinkel, J., Natalin, P., Stark, A., Brennecke, J., Cohen, S.M., and Izaurralde, E. (2006). Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in Drosophila melanogaster. Mol Cell Biol *26*, 2965-2975.
Sachs, M.S., Wang, Z., Gaba, A., Fang, P., Belk, J., Ganesan, R., Amrani, N., and Jacobson, A. (2002). Toeprint analysis of the positioning of translation apparatus components at initiation and termination codons of fungal mRNAs. Methods *26*, 105-114.
Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58-63.
Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-862.
Wu, L., Fan, J., and Belasco, J.G. (2006). MicroRNAs direct rapid deadenylation of mRNA. Proc Natl Acad Sci U S A *103*, 4034-4039.

Figure 1

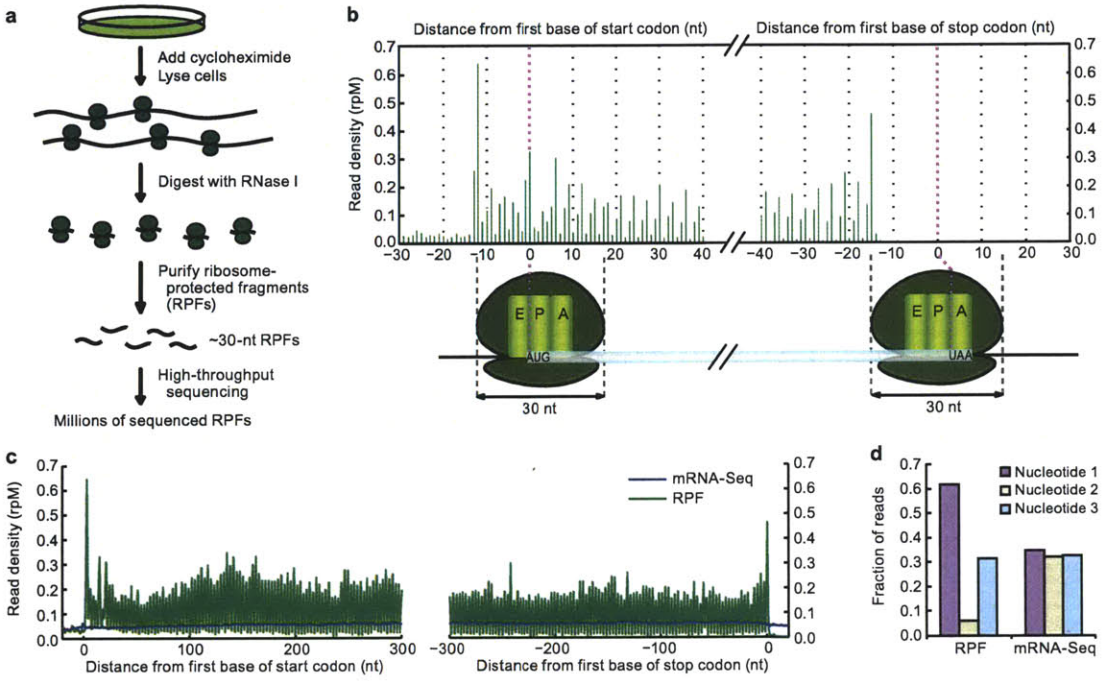

a
Add cycloheximide
Lyse cells
Digest with RNase I
Purify ribosome-protected fragments (RPFs)
~30-nt RPFs
High-throughput sequencing
Millions of sequenced RPFs
b
Distance from first base of start codon (nt)
Distance from first base of stop codon (nt)
Read density (rpM)
E P A
AUG
UAA
30 nt
30 nt
c
Read density (rpM)
mRNA-Seq
RPF
Distance from first base of start codon (nt)
Distance from first base of stop codon (nt)
d
Fraction of reads
Nucleotide 1
Nucleotide 2
Nucleotide 3
RPF
mRNA-Seq

Figure 2

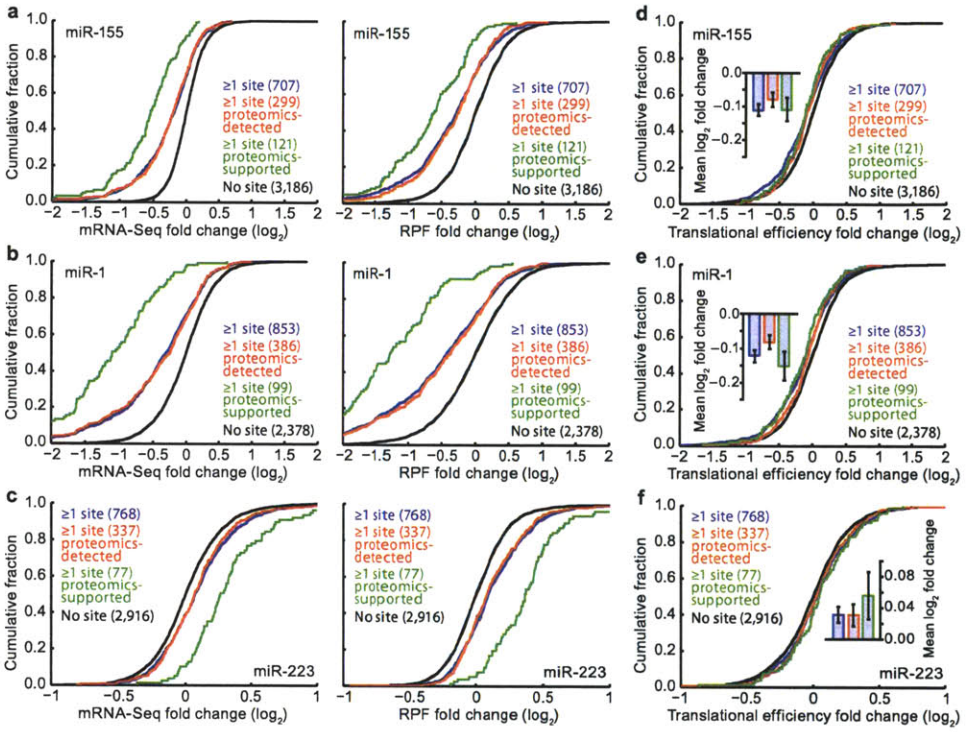

a
miR-155
Cumulative fraction
≥1 site (707)
≥1 site (299) proteomics-detected
≥1 site (121) proteomics-supported
No site (3,186)
mRNA-Seq fold change (log2)
RPF fold change (log2)
d
Mean log2 fold change
Translational efficiency fold change (log2)
b
miR-1
≥1 site (853)
≥1 site (386) proteomics-detected
≥1 site (99) proteomics-supported
No site (2,378)
e
c
miR-223
≥1 site (768)
≥1 site (337) proteomics-detected
≥1 site (77) proteomics-supported
No site (2,916)
f

Figure 3

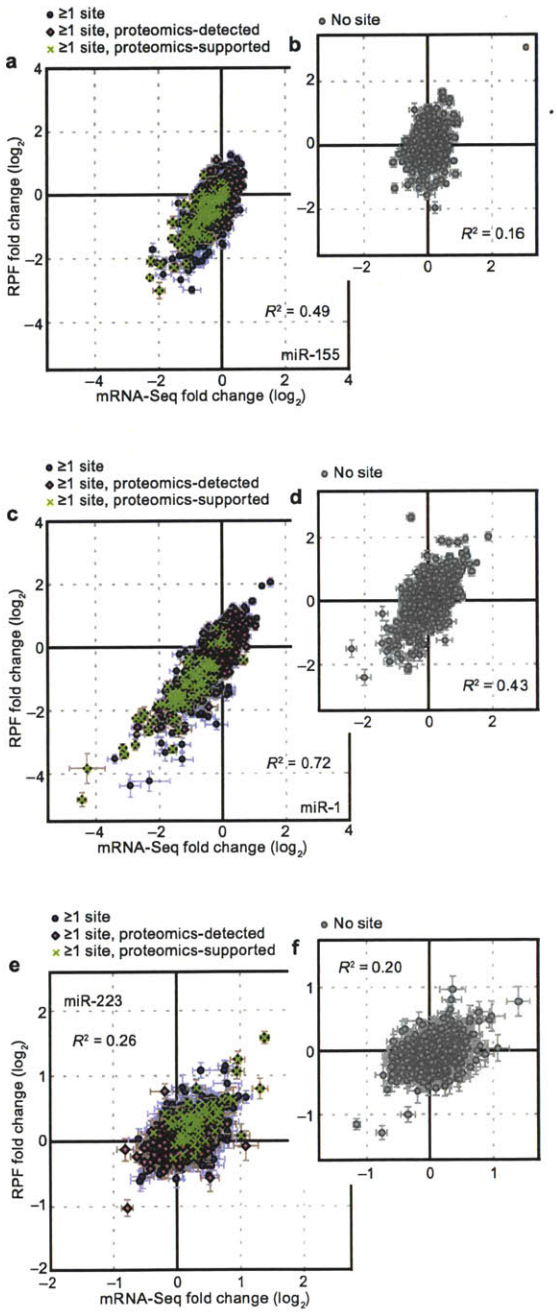
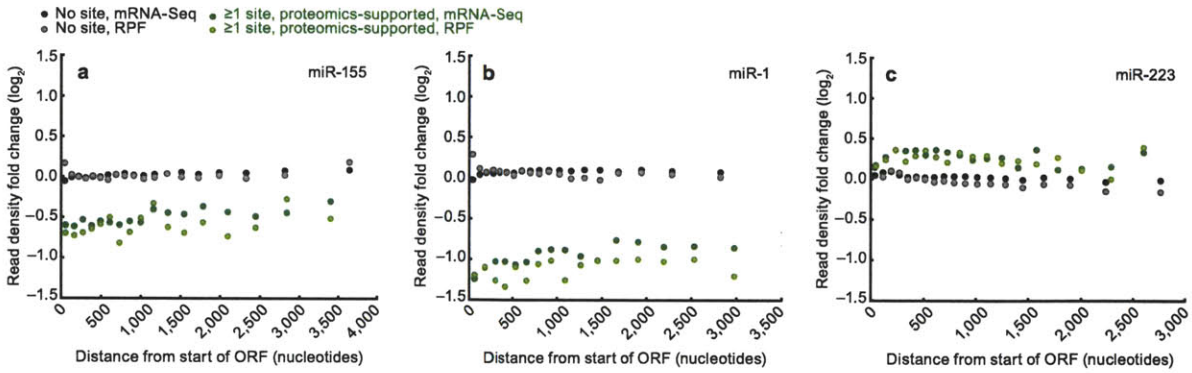

≥1 site
≥1 site, proteomics-detected
≥1 site, proteomics-supported
No site
a
b
c
d
e
f
RPF fold change (log2)
mRNA-Seq fold change (log2)
R² = 0.49
miR-155
R² = 0.16
R² = 0.72
miR-1
R² = 0.43
miR-223
R² = 0.26
R² = 0.20

Figure 4


No site, mRNA-Seq
No site, RPF
≥1 site, proteomics-supported, mRNA-Seq
≥1 site, proteomics-supported, RPF
a
miR-155
b
miR-1
c
miR-223
Read density fold change (log2)
Distance from start of ORF (nucleotides)

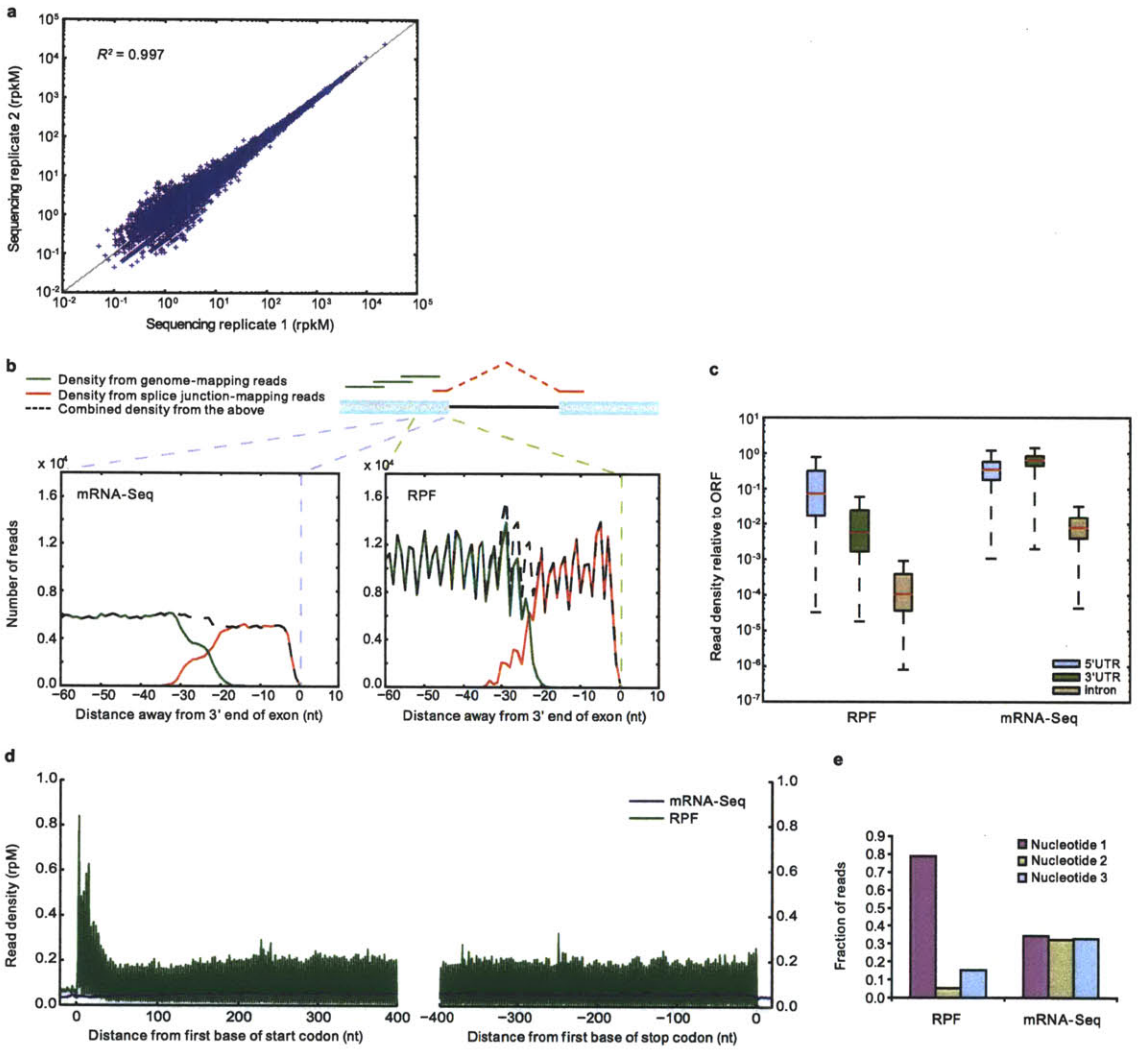
Supplemental materials

Supplementary Table 1. Alignment statistics for sequencing reads.

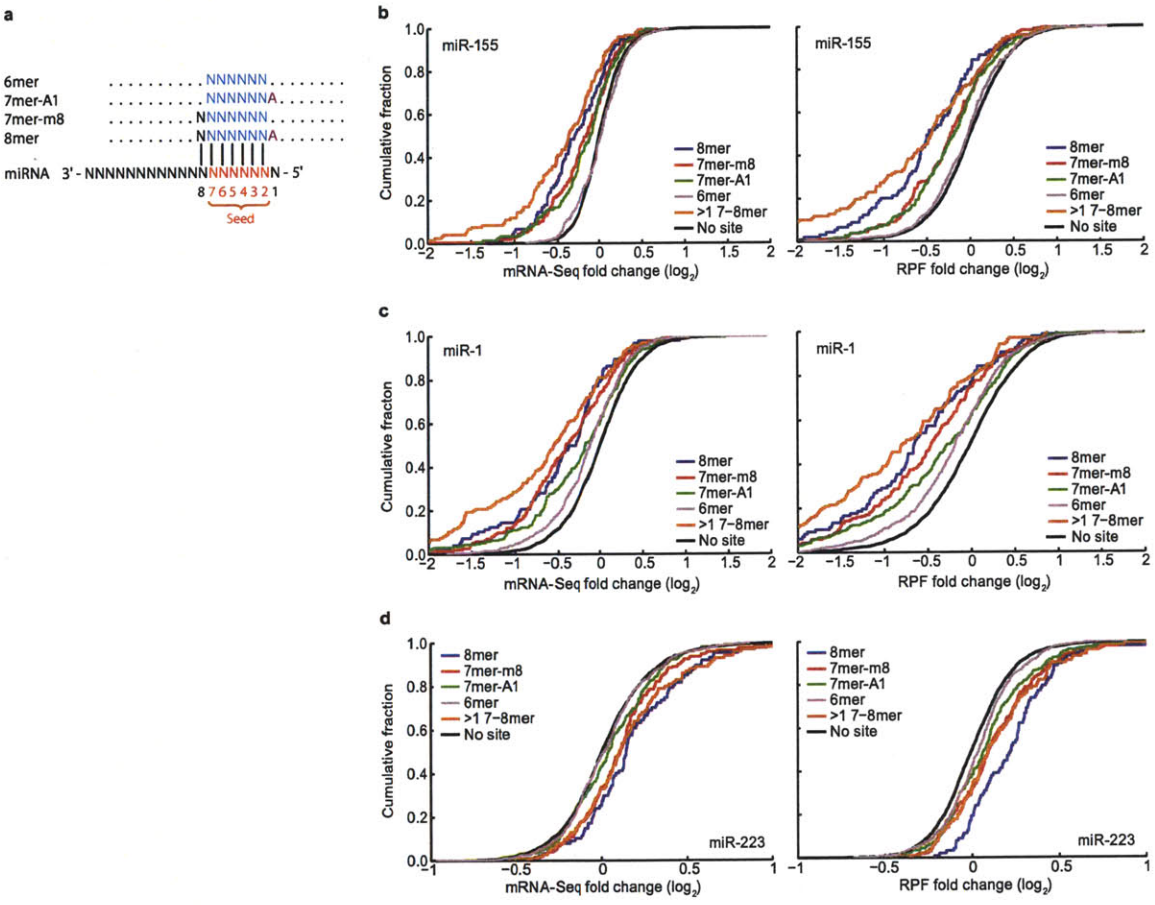
	HeLa mRNA-Seq, 32 h			HeLa Ribosome Profiling, 32 h			Neutrophil mRNA-Seq		Neut. Ribosome Profiling	
	Mock	miR-155	miR-1	Mock	miR-155	miR-1	WT	*mir-223*$^{-/Y}$	WT	*mir-223*$^{-/Y}$
Raw number of reads	11,441,416	14,392,817	13,505,056	18,029,685	17,454,122	17,643,682	17,303,502	16,280,254	16,418,487	18,072,590
Uniquely mapping reads										
To genome (first-phase)	5,722,434	7,447,242	7,179,753	8,533,140	8,975,586	8,902,916	8,362,287	7,806,247	10,239,553	11,080,166
To reference transcript database (second-phase)	403,577	499,784	487,120	758,639	753,094	704,680	593,777	584,578	788,191	869,904
Total mapping uniquely	6,126,011	7,947,026	7,666,873	9,291,779	9,728,680	9,607,596	8,956,064	8,390,825	11,027,744	11,950,070
Percentage of raw reads mapping uniquely	53.5	55.2	56.8	51.5	55.7	54.5	51.8	51.5	67.2	66.1
rRNA unique matches	748,043	968,296	803,583	3,114,740	3,641,076	3,792,787	365,308	439,258	4,633,841	4,831,036
Percentage rRNA unique matches	12.2	12.2	10.5	33.5	37.4	39.5	4.1	5.2	42.0	40.4
Unambiguous, unique gene-mapping reads										
Total mapped reads (library size)	4,606,993	6,019,691	5,910,777	5,439,248	5,342,830	5,102,548	7,652,330	7,070,727	6,002,118	6,688,548
Percentage of all unique matches	75.2	75.8	77.1	58.5	54.9	53.1	85.4	84.3	54.4	56.0
Number of reads mapping to exons	4,095,694	5,353,615	5,249,653	5,379,200	5,279,217	5,037,521	6,542,225	6,162,103	5,945,873	6,627,659
Number of reads mapping to coding exons	2,762,359	3,612,248	3,493,711	5,136,392	5,011,176	4,784,714	4,500,718	4,275,842	5,562,427	6,200,226
Number of reads mapping to introns	511,299	666,076	661,124	60,048	63,613	65,027	1,110,105	908,624	56,245	60,889
Percentage coding exon reads, out of all reads mapping to exons	67.5	67.5	66.6	95.5	94.9	95.0	68.8	69.4	93.6	93.6

Supplementary Table 2. Alignment statistics for sequencing reads, for samples collected at 12 h post-transfection.

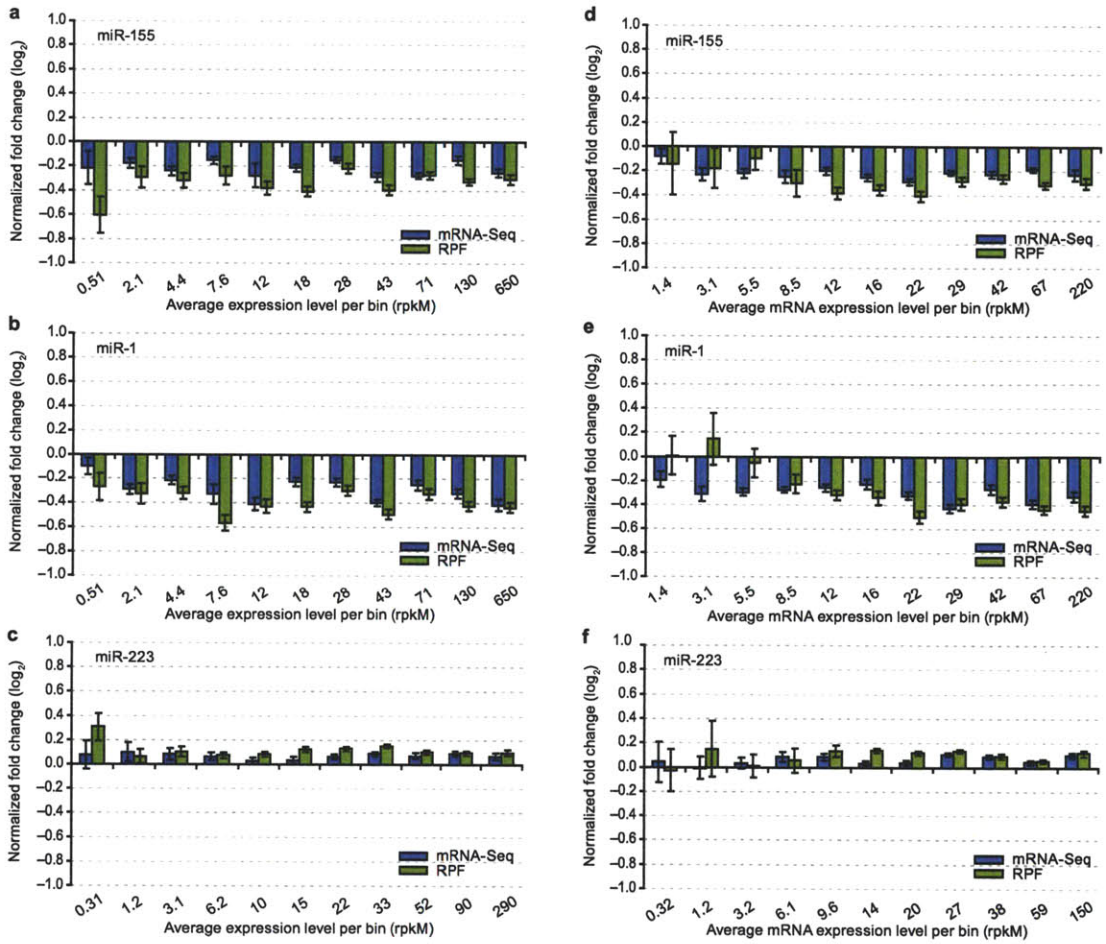
	HeLa mRNA-Seq, 12 h			HeLa Ribosome Profiling, 12 h		
	Mock	miR-155	miR-1	Mock	miR-155	miR-1
Raw number of reads	21,623,454	21,388,074	21,681,406	16,793,006	16,792,228	15,741,337
Uniquely mapping reads						
To genome (first-phase)	10,743,704	11,232,931	11,957,592	7,882,398	7,923,788	6,669,746
To reference transcript database (second-phase)	708,950	753,144	771,463	722,046	912,444	693,775
Total mapping uniquely	11,452,654	11,986,075	12,729,055	8,604,444	8,836,232	7,363,521
Percentage of raw reads mapping uniquely	53.0	56.0	58.7	51.2	52.6	46.8
rRNA unique matches	1,172,686	1,326,092	1,375,560	2,984,080	1,802,419	1,851,953
Percentage rRNA unique matches	10.9	11.8	11.5	37.9	22.7	27.8
Unambiguous, unique gene-mapping reads						
Total mapped reads (library size)	8,746,696	9,147,411	9,706,156	4,999,224	6,296,140	4,927,498
Percentage of all unique matches	76.4	76.3	76.3	58.1	71.3	66.9
Number of reads mapping to exons	7,524,106	8,018,834	8,362,731	4,940,460	6,217,253	4,864,783
Number of reads mapping to coding exons	4,882,314	5,340,751	5,544,369	4,712,453	5,908,841	4,590,167
Number of reads mapping to introns	1,222,590	1,128,577	1,343,425	58,764	78,887	62,715
Percentage coding exon reads, out of all reads mapping to exons	64.9	66.6	66.3	95.4	95.0	94.4



Supplementary Figure 1. Ribosome profiling in mammalian cells captured features of translation. **a**, Correspondence between gene expression quantified in two different sequencing replicates of the same library. Expression values are in terms of reads per million reads mapped to genes per kilobase coding exon model (rpkM). The R^2 derived from Pearson's correlation is indicated. **b**, Density of mRNA-Seq tags and RPFs near the 3' ends of exons, which indicates that splice isoforms used in our reference transcript database were representative of the transcriptome. Plotted are 5' termini of mRNA-Seq tags (*left*) or RPFs (*right*), for genome-mapping reads (green solid line), splice junction-mapping reads (red solid line) and the combined read density (black dashed line). Composite data are shown for all exons ≥100 nucleotides (nt) in length. Athough reads coming from the rest of the exon can map to the genome irrespective of which splice isoforms were expressed, reads spanning the splice junctions might not have sufficient length on the 5' side of the junction, to be confidently mapped by the mapping program. Hence, if a gene were expressed in multiple isoforms in the cell or if we picked the wrong isoform, reads coming from the 3' ends of certain exons of the correct isoforms might not be mapped. This was, however, not a major problem as the read density obtained for this region overall was >90% that of the rest of the exon. The jagged pattern observed in the plot for RPFs was explained by most exons being in the same reading frame (together with the codon nucleotide preference observed with ribosome profiling, e.g., Fig. 1d and Supplementary Fig. 1e) **c**, Expression levels from different regions of mRNAs, quantified by RPFs and mRNA-Seq tags. Expression values from each indicated region in a transcript was normalized to that in the associated ORF and plotted as relative density, as quantified by RPFs and mRNA-Seq tags. **d**, Density of RPFs and mRNA-Seq tags near the beginnings and ends of ORFs in mouse neutrophils. RPF and mRNA-Seq density are plotted as in Fig. 1c. Composite data are shown for ≥ 600-nucleotide ORFs that passed our threshold for quantification (≥100 RPFs and ≥100 mRNA-Seq tags). **e**, Fraction of RPFs and mRNA-Seq tags mapping to each of the three codon nucleotides in panel **d**. Our samples exhibited small differences in the ratio of reads mapping to codon nucleotides 3 and 1, which we attribute to the extent of RNase I digestion, with more reads at codon nucleotide 3 correlated with less thorough digestion (as evaluated on sucrose gradient profiles). Our mouse neutrophil samples were more extensively digested than were the HeLa samples, and the reads exhibited an even stronger (80%) preference for codon nucleotide 1.

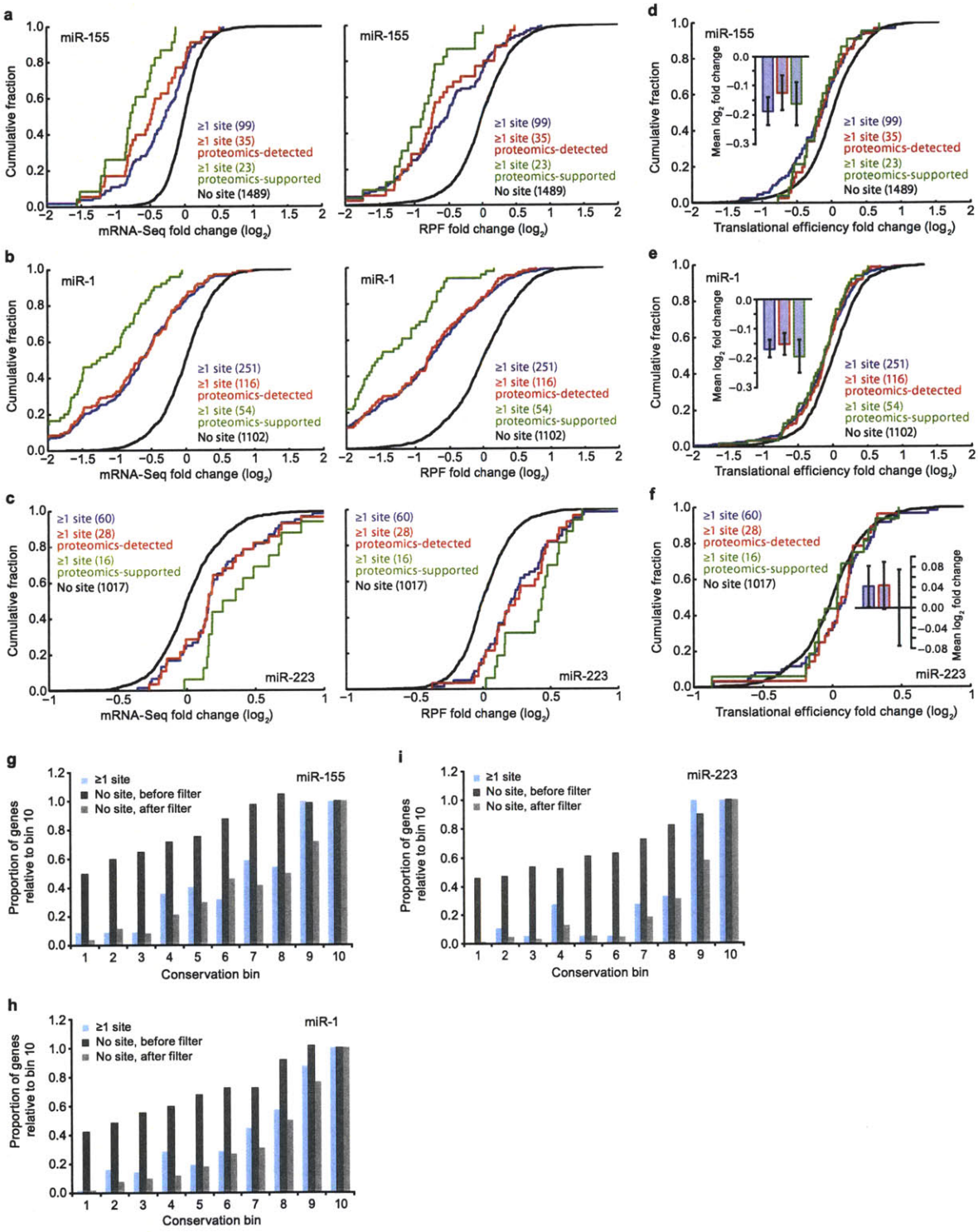


Supplementary Figure 2. Analyses of miRNA-dependent changes, which showed that ribosome profiling, like mRNA profiling, captured the characteristic hierarchy of site efficacies. **a**, Different types of seed-matched sites (Bartel, 2009). **b**, Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right*) after introducing miR-155. Plotted are distributions for genes with either a single 3'UTR site of the indicated type, multiple 7–8mer 3'UTR sites, or no site. Only genes that passed our threshold for quantification (≥100 RPFs and ≥100 mRNA-Seq tags) were considered. **c**, Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right*) after introducing miR-1. Otherwise, as in panel **b**. **d**, Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right*) after deleting *mir-223*. Otherwise, as in panel **b**.



Supplementary Figure 3. MicroRNA targets expressed at different levels were similarly repressed. **a**, Changes in RPFs and mRNA-Seq tags from mRNAs with cognate sites, after introducing miR-155 into HeLa cells. Genes were binned in groups of 1000 based on expression, as measured by RPF density in mock-transfected cells. The average RPF expression density for each bin is indicated in terms of reads per million reads mapped to genes per kilobase coding exon model (rpkM). Plotted for each bin are the normalized aggregate fold changes in RPFs (green) and mRNA-Seq tags (blue). In each case, the normalized aggregate fold change was derived by dividing the aggregate fold change for genes with 3'UTR sites by that for genes without sites anywhere in their mRNA. To estimate the average aggregate fold change value for genes without sites, 100 control cohorts were generated by randomly selecting genes without sites to replace genes with sites. The mean fold change for these 100 cohorts was then used for normalization in each bin. Error bars represent the standard deviations of each of these 100 cohorts. **b**, As in panel **a**, but plotting repression of mRNAs with miR-1 3'UTR sites after introducing miR-1. **c**, As in panel **a**, but plotting derepression of mRNAs with miR-223 3'UTR sites, comparing samples from neutrophils derived from *mir-223* knockout mice to those from neutrophils derived from wild-type mice. Genes were binned according to their expression levels in the KO dataset. The top 80 most highly expressed genes were excluded from this analysis because their high expression levels (otherwise comprising >60% of the total RPFs in the last bin) led to high variability for the control cohorts of the last bin. **d**, As in panel **a**, but genes were binned in groups of 1000 based on expression as measured by mRNA-Seq density (rpkM) rather than RPF density. **e**, As in panel **b**, but genes were binned in groups of 1000 based on expression as measured by mRNA-Seq density (rpkM). **f**, As in panel **c**, but genes were binned in groups of 1000 based on expression as measured by mRNA-Seq density (rpkM). The top 80 most highly expressed genes were again excluded from the analysis (otherwise comprising >56% of the total mRNA-Seq tags in the last bin).

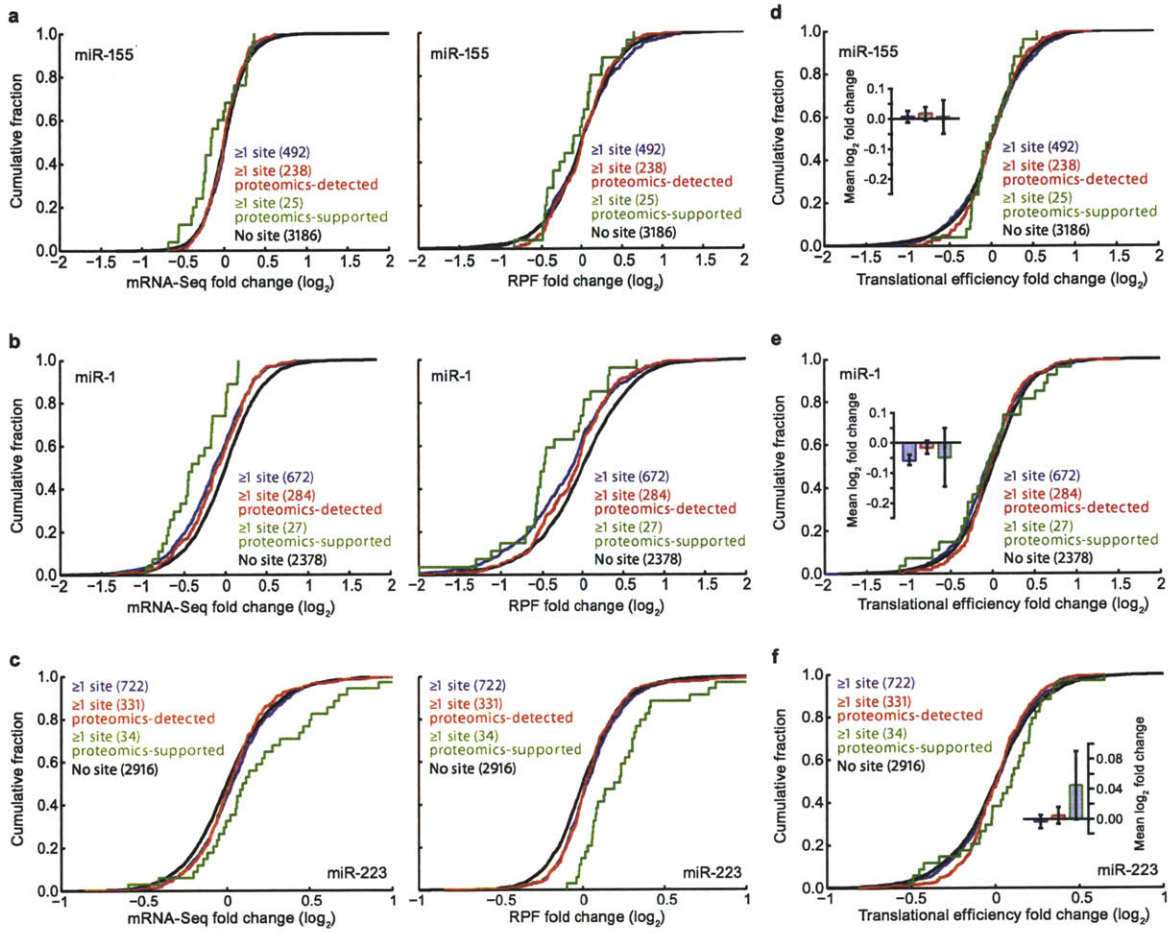
To examine a potential relationship between gene expression and miRNA-mediated repression, genes were grouped by expression level, and then for each bin the results for messages with at least one 3'UTR site were compared to those without a site anywhere in the mRNA. For example, in the miR-155 experiment, genes with uniquely mapping RPFs were assigned to bins, with 1000 distinct genes per bin, based on the density of RPFs in mock-transfected HeLa cells (panel **a**). Because genes in each bin were analyzed in aggregate, a read threshold for individual genes was not required, allowing inclusion of data from 11,000 distinct genes, which ranged broadly in expression (more than 1000-fold difference between the first and last bins). Except for one outlier bin, messages with miR-155 sites underwent repression of similar magnitude regardless of their expression level (panel **a**). The same was true in the miR-1 experiment (panel **b**). When we examined endogenous miRNA:target interactions, as detected by monitoring derepression ($\log_2$-fold increases rather than decreases) in *mir-223* KO cells, analogous results were obtained (panel **c**). Although the RPF change for the bin with the lowest-expressed genes appeared greater than that for other bins in the miR-155 and miR-223 experiments, this change was accompanied by larger variability (note error bars) and was not observed in neighboring bins, nor when the analysis was repeated using mRNA-Seq tags to group the genes based on mRNA expression (panels **d**–**f**). Moreover, in the miR-1 experiment the RPF change for this bin was no greater than that for other bins (panel **b**), further supporting the conclusion that miRNAs do not repress their lowly expressed targets more potently than they do their moderately expressed targets.



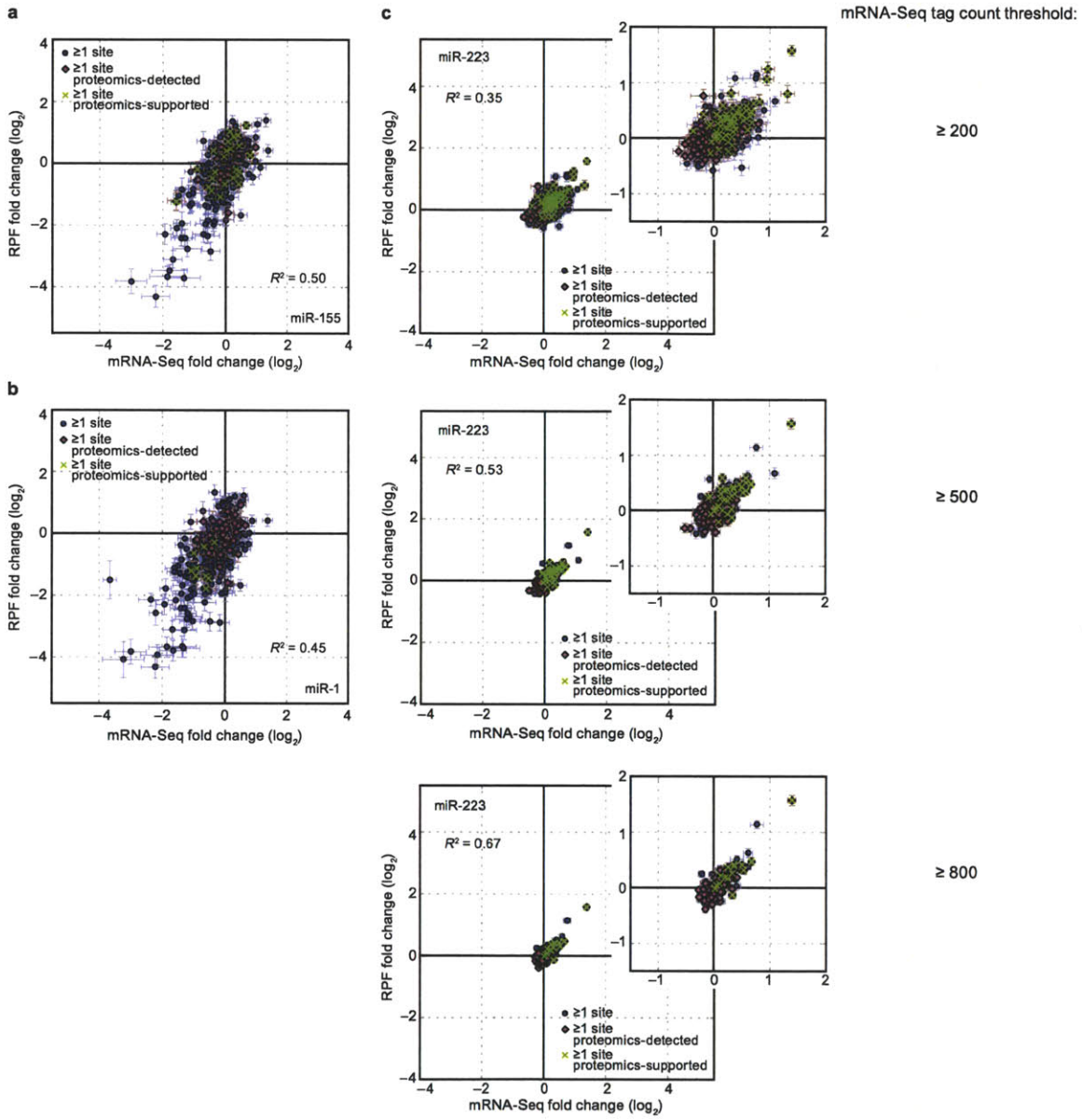
Supplementary Figure 4. MicroRNA targets with conserved 3'UTR sites were repressed mostly through mRNA destabilization. **a**, Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right*) after introducing miR-155, for genes with conserved sites to miR-155. Plotted are distributions for the genes with ≥1 miR-155 conserved 3'UTR site (blue), the subset of these genes detected in the pSILAC experiment (proteomics-detected, red), the subset of the proteomics-detected genes with proteins responding with $\log_2$-fold change ≤ –0.3 (proteomics-supported, green), and similarly-conserved control genes, which lacked miR-155 sites throughout their mRNAs (no site, black). The number of genes in each category is indicated in parentheses. Only genes that passed our expression thresholds (≥100 RPFs and ≥100 mRNA-Seq tags) were considered. Note that the "no site" distribution comprised genes selected such that their 3'UTRs were conserved at similar levels as the 3'UTRs with conserved sites (discussed below), hence the smaller number of genes compared to the "no site" distribution in Fig. 2a. **b**, Cumulative distributions of

mRNA-Seq changes (*left*) and RPF changes (*right*) after introducing miR-1. Otherwise, as in panel **a**. **c**, Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right)* after deleting *mir-223*. Otherwise, as in panel **a**, with proteomics-supported genes referring to genes with proteins that responded with $\log_2$-fold change ≥ 0.3 in the SILAC experiment. **d**, Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after introducing miR-155. For each gene, the translational efficiency change was calculated by normalizing the RPF change by the mRNA-Seq change. The cumulative distributions of translational efficiency changes were then plotted, as in panel **a**. For each distribution, the mean $\log_2$-fold change ($\pm$ standard error) is shown in the inset. **e**, Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after introducing miR-1. Otherwise, as in panel **d**. **f**, Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after deleting *mir-223*. Otherwise, as in panel **d**. **g**, Distribution of quantified genes in the miR-155 experiment across different conservation bins. Quantified genes with conserved sites to miR-155 (blue bars) are represented in terms of their proportion across the different conservation bins (Friedman et al., 2009), with the fraction in bin 10 (most highly conserved) set to 1.0. Similarly, the proportions of quantified genes without sites are shown before (dark grey bars) and after (light grey bars) controlling for conservation levels, as described in the following paragraph. Genes in the latter group made up the "no site" distribution plotted in panels **a** and **d**. **h**, Distribution of quantified genes in the miR-1 experiment across different conservation bins. Otherwise, as in panel **g**. **i**, Distribution of quantified genes in the miR-223 experiment across different conservation bins. Otherwise, as in panel **g**.

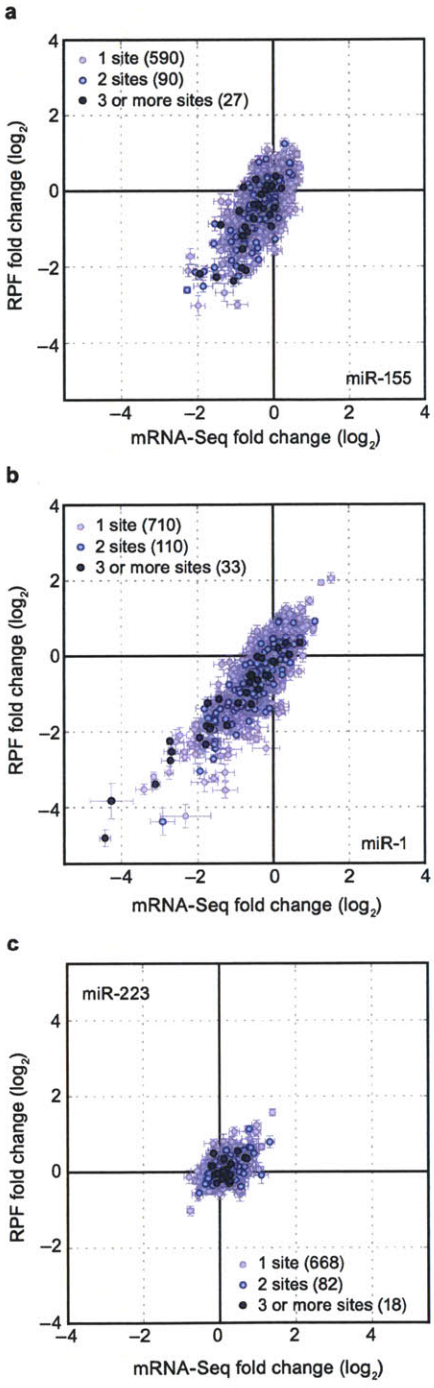
Due to concern that 3'UTRs with conserved sites were expected to be more conserved than those without sites, which might influence our analyses of conserved targets, we controlled for conservation levels using data from Friedman et al. (2009), who divide genes into 10 bins based on 3'UTR conservation levels. We first determined the number of genes with sites in each of these bins. Then genes without sites were sampled from each bin until the population of genes without sites reached the same proportions across the different bins as the population of genes with sites. With this filter, genes that formed the "no site" distribution had 3'UTR conservation levels matching that of the conserved targets. For all three experiments, controlling for 3'UTR conservation did not significantly change the cumulative distribution plots for the no-site subsets of genes ($P > 0.24$, K–S test).



Supplementary Figure 5. MicroRNA targets with sites only in their ORFs were weakly repressed, mostly through mRNA destabilization. **a**, Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right*) after introducing miR-155. Plotted are distributions for the genes with ≥1 miR-155 ORF site (blue), the subset of these genes detected in the pSILAC experiment (proteomics-detected, red), the subset of the proteomics-detected genes with proteins responding with $\log_2$-fold change ≤ –0.3 (proteomics-supported, green), and the control genes, which lacked miR-155 sites throughout their mRNAs (no site, black). The number of genes in each category is indicated in parentheses. Only genes that passed our expression thresholds (≥100 RPFs and ≥100 mRNA-Seq tags) were considered, and genes with a miR-155 3'UTR site (6-nt seed match) were excluded. **b**, Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right*) after introducing miR-1. Otherwise, as in panel **a**. **c**, Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right*) after deleting *mir-223*. Otherwise, as in panel **a**, with proteomics-supported genes referring to genes with proteins that responded with $\log_2$-fold change ≥ 0.3 in the SILAC experiment. **d**, Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after introducing miR-155. For each gene, the translational efficiency change was calculated by normalizing the RPF change by the mRNA-Seq change. The cumulative distributions of translational efficiency changes were then plotted, as in panel **a**. For each distribution, the mean $\log_2$-fold change (± standard error) is shown in the inset. **e**, Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after introducing miR-1. Otherwise, as in panel **d**. **f**, Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after deleting *mir-223*. Otherwise, as in panel **d**.

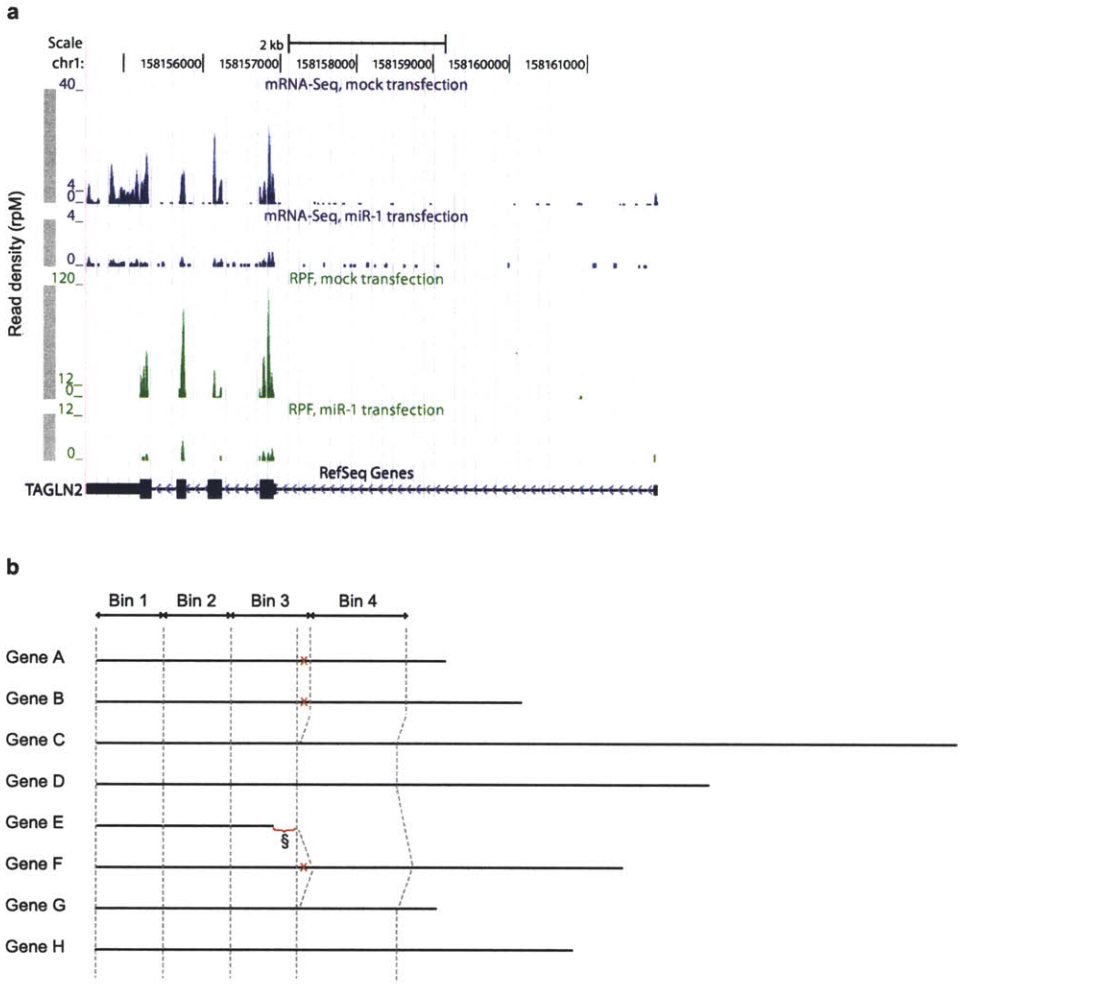


Supplementary Figure 6. Correspondence between ribosome changes and mRNA changes from miRNA targeting when using alternative quantification thresholds. **a**, Correspondence between ribosome (RPF) and mRNA (mRNA-Seq) changes after introducing miR-155, for genes that were only considered after relaxing the quantification thresholds from ≥100 reads each in the mRNA-Seq and RPF libraries (mock transfection) to ≥44 mRNA-Seq tags and ≥80 RPF reads (blue circles). Proteomics-detected targets and proteomics-supported targets are highlighted (pink diamonds and green crosses, respectively). Expected standard deviations (error bars) were calculated based on the number of reads obtained per gene and assuming random counting statistics. The R^2 derived from Pearson's correlation of all data is indicated. **b**, Correspondence between ribosome and mRNA changes after introducing miR-1, for genes that were only considered after relaxing the quantification thresholds from ≥100 reads each in the mRNA-Seq and RPF libraries (mock transfection) to ≥44 mRNA-Seq tags and ≥80 RPF reads. Otherwise, as in panel **a**. **c**, Correspondence between ribosome and mRNA changes after deleting *mir-223*, for genes considered at increasing stringency of mRNA-Seq quantification thresholds. The RPF read threshold was kept constant at ≥100 reads in the KO sample, while the threshold for the mRNA-Seq KO library was increased from ≥100 reads in Fig. 3e to ≥200 reads, ≥500 reads and ≥800 reads. Otherwise, as in panel **a**. Insets provide closer views of the corresponding sets of data.

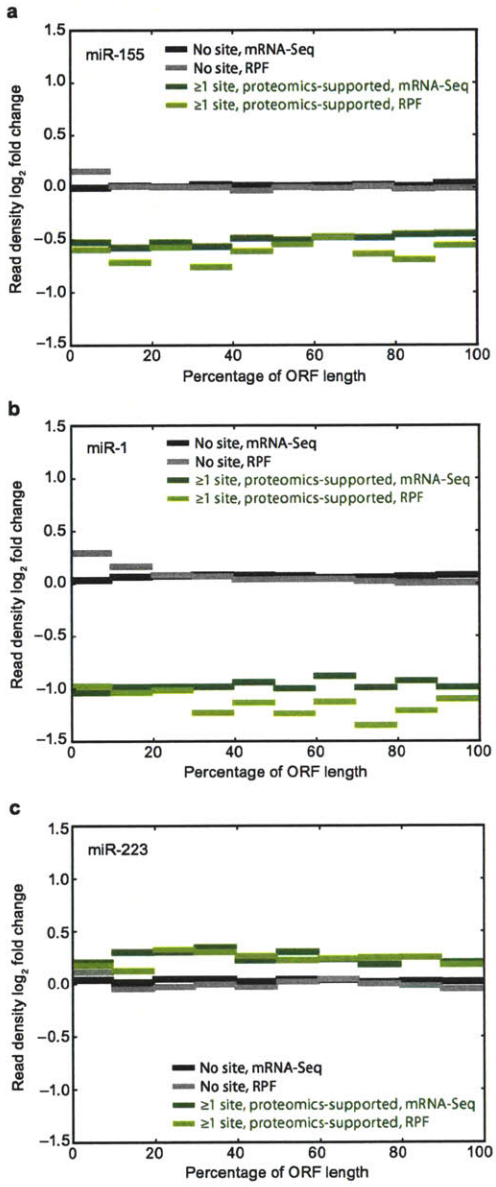


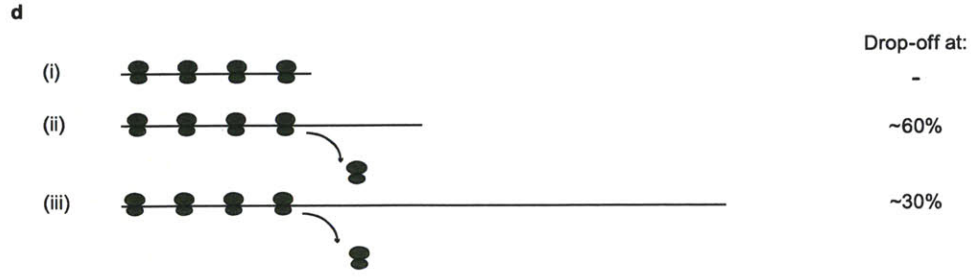
Supplementary Figure 7. Genes with multiple sites to the cognate miRNA were repressed in a manner similar to that of genes with single sites. **a**, Correspondence between ribosome (RPF) and mRNA (mRNA-Seq) changes after introducing miR-155, plotting data for the 707 quantified genes with at least one miR-155 3′UTR site. Genes with two sites and genes with at least three sites are highlighted (light blue circles and dark blue circles, respectively). Expected standard deviations (error bars) were calculated based on the number of reads obtained per gene and assuming random counting statistics. **b**, Correspondence between ribosome and mRNA changes after introducing miR-1, plotting data for the 853 quantified genes with at least one miR-1 3′UTR site. Otherwise, as in panel **a**. **c**, Correspondence between ribosome and mRNA changes after deleting *mir-223*, plotting data for the 768 quantified genes with at least one miR-223 3′UTR site. Otherwise, as in panel **a**.

Linear regression lines for each category of genes were calculated and tested against each other to see whether genes with different number of 3′UTR sites to the cognate miRNA responded differently. Of these comparisons, only two yielded statistically significant differences. In the miR-155 experiment, the regression slope of the genes with one site was different from that of the genes with two sites [$P = 0.04$, two-tailed Analysis of Covariance (ANCOVA) test]. In the miR-1 experiment, the regression slope of the genes with two sites was different from that of the genes with at least 3 sites ($P = 0.02$), but the difference in slope was in the opposite direction as observed for the greater number of sites in the miR-155 experiment. Because a consistent difference was not observed in any two datasets, we conclude that genes with multiple 3′UTR sites to the cognate miRNA were repressed in a manner similar to that of genes with single sites, i.e., mostly through mRNA destabilization.



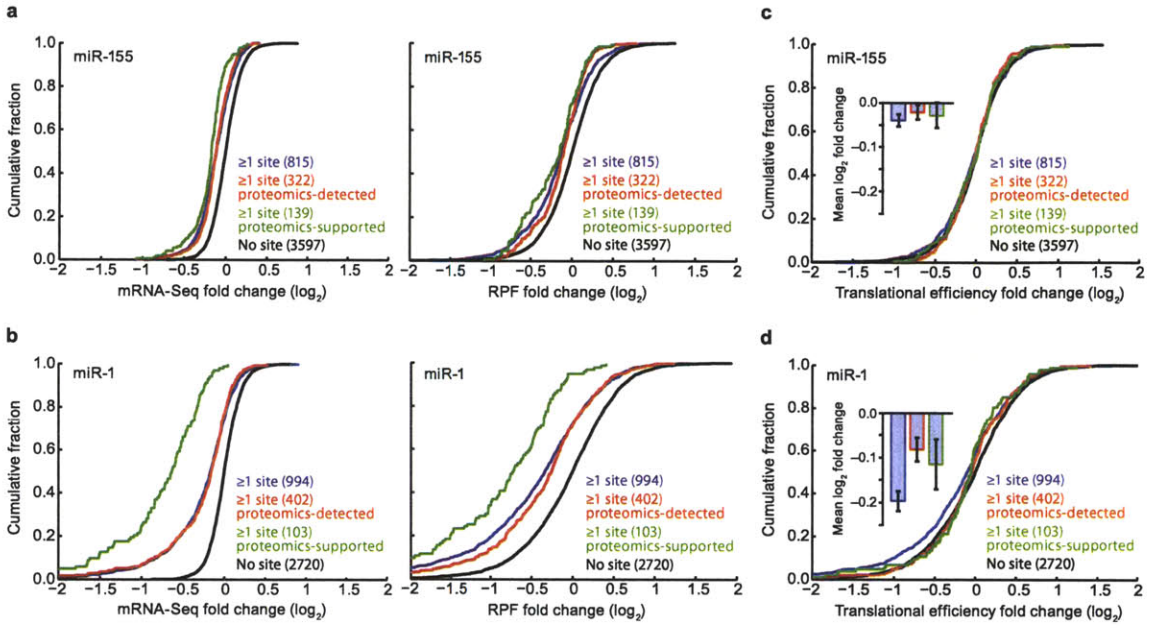
Supplementary Figure 8. Probing repression along the length of the mRNA. **a**, Read density from a gene with high sequence coverage. Normalized read density (mRNA-Seq, blue and RPF, green) for *TAGLN2*, a miR-1 target, is shown for the mock and miR-1 transfections. Note the ten-fold difference in the range of the y-axes between the coverage plots for the mock and miR-1 transfections. **b**, Schematic diagram of the binning strategy used in Fig. 4. All open reading frames within each category were lined up from the first nucleotide. All bases available were then assigned into 20 bins by drawing the same number of bases from each gene, for each bin. If a shortfall of bases resulted from reaching the ends of shorter genes (e.g., a shortfall of 3 nucleotides in Bin 3), the shortfall was alleviated by extending randomly selected representatives of the remaining genes (red x). The formation of subsequent bins then started from the remaining unassigned bases, as illustrated by the shifted frame. Later bins received a higher contribution from each remaining gene (as illustrated by the increased width in Bin 4), because a smaller number of genes remained after all bases in the shorter genes (represented by Gene E) had been assigned. This explains why the median bases plotted on the x-axes in Fig. 4 become further apart with increasing distance into the composite ORF. With this approach, every bin has the same number of nucleotides assigned to it (except the very last bin, which was discarded) and each successive bin represents a distance further away from the start of the composite ORF. To prevent spurious changes of just a few genes from dominating, bins were discarded if their read contribution derived from fewer than 20 genes. Because the number of genes with sites is smaller than that without sites, the 20-gene limit came at an earlier bin for the genes with sites, which explains why in Fig. 4 the plot for genes with sites has fewer bins than that for genes without sites.



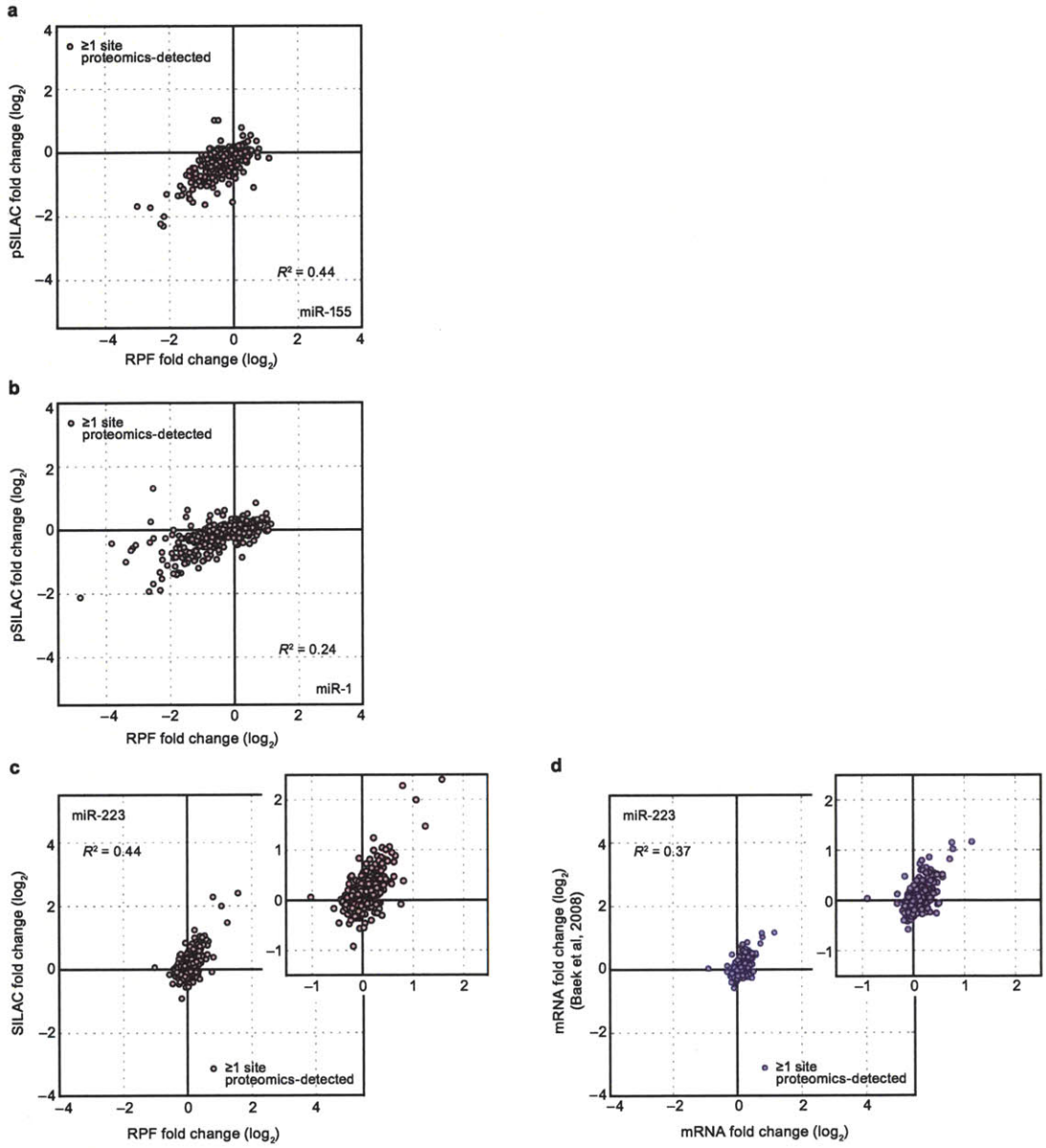


Supplementary Figure 9. Ribosome and mRNA changes were uniform along the percentage length of the ORFs. **a**, Ribosome and mRNA changes along the percentage length of the ORFs after introducing miR-155. The ORF of each quantified gene was divided into ten bins based on ORF length. Fold changes in RPFs and mRNA-Seq tags mapping to each bin were then plotted. Changes in RPFs and mRNA-Seq tags for mRNAs with no site (grey and black, respectively) and for proteomic-supported targets (light and dark green, respectively) are shown. **b**, Ribosome and mRNA changes along the percentage length of the ORFs after introducing miR-1. Otherwise, as in panel **a**. **c**, Ribosome and mRNA changes along the percentage length of the ORFs after deleting *mir-223*. Otherwise, as in panel **a**. **d**, Schematic diagram illustrating how a percentage-based measure could obscure drop-off effects. In this scenario, ribosomes start to drop off after traversing a certain length of the ORF. Based on the different lengths of the different ORFs, a drop-off at a similar absolute distance would occur at different percentage ORF lengths. In the extreme case of very short genes (i), no drop-off would occur and the lack of change at the ends of these ORFs could thereby obscure the drop-off effect from longer genes.

Due to the caveat presented in panel **d**, we adopted the alternative binning strategy used in Fig. 4, which makes use of a distance-based measure (Supplementary Fig. 8b). We note that for the miR-1 experiment, the percentage-based plot could suggest drop-off (panel **b**). However, if this were indeed the mechanism, this effect should become more obvious in the distance-based plot (Fig. 4b). The apparent incongruence of the two plots for the miR-1 experiment could be due to a few short genes with fewer reads towards the 3' ends of their ORFs having high sequence coverage and thus dominating changes in the percentage-based plot. The distance-based plot would not be confounded by this effect because shorter genes can only contribute to the early bins. To prevent the reciprocal effect of longer genes dominating the later bins and thus biasing the distance-based plots, we only plotted the bins which had read contribution from at least 20 genes in Fig. 4. Thus, for us to observe a drop-off effect in the distance-based plots, ribosome drop-off must be a mechanism that is sufficiently general to affect more than a few genes and genes of different length ranges. Because we do not see this effect, we conclude that for most targets ribosome drop-off is likely not a major component of the repression mechanism.



Supplementary Figure 10. At an earlier time point post-transfection most of the downregulation mediated by miRNAs was mostly through mRNA destabilization. **a**, Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right*) after introducing miR-155. Plotted are distributions for the genes with ≥1 miR-155 3'UTR site (blue), the subset of these genes detected in the pSILAC experiment (proteomics-detected, red), the subset of the proteomics-detected genes with proteins responding with $\log_2$-fold change ≤ –0.3 (proteomics-supported, green), and the control genes, which lacked miR-155 sites throughout their mRNAs (no site, black). The number of genes in each category is indicated in parentheses. Only genes that passed our expression thresholds (≥100 RPFs and ≥100 mRNA-Seq tags) were considered. **b**, Cumulative distributions of mRNA-Seq changes (*left*) and RPF changes (*right*) after introducing miR-1. Otherwise, as in panel **a**. **c**, Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after introducing miR-155. For each gene, the translational efficiency change was calculated by normalizing the RPF change by the mRNA-Seq change. The cumulative distributions of translational efficiency changes were then plotted, as in panel **a**. For each distribution, the mean $\log_2$-fold change (± standard error) is shown in the inset. **d**, Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after introducing miR-1. Otherwise, as in panel **c**.



Supplementary Figure 11. Correspondence between ribosome changes, as measured by ribosome profiling, and changes in protein output, as measured by previous proteomics experiments. **a**, Correspondence between ribosome (RPF) and protein changes after introducing miR-155, plotting data for the 299 genes that had at least one miR-155 3'UTR site and were detected in the pSILAC experiment (proteomics-detected). The R^2 derived from Pearson's correlation is indicated. **b**, Correspondence between ribosome and protein changes after introducing miR-1, plotting data for the 386 genes that had at least one miR-1 3'UTR site and were detected in the pSILAC experiment. Otherwise, as in panel **a**. **c**, Correspondence between ribosome and protein changes after deleting *mir-223*, plotting data for the 337 genes that had at least one miR-223 3'UTR site and were detected in the SILAC experiment. Otherwise, as in panel **a**. The inset provides a closer view of the same data. **d**, Correspondence between mRNA changes in the Baek et al. (2008) experiment and the mRNA changes in the current experiment. Both axes plot fold changes measured on mRNA arrays for the 310 genes that 1) had at least one miR-223 3'UTR site, 2) were detected in the SILAC experiment, and 3) were assigned the identical transcript isoform in the two experiments.

In the experiments that ectopically introduced miRNAs, the ribosome changes tended to be larger than the protein changes (panels **a** and **b**). This difference might reflect different efficiencies of the miRNA transfections, which were performed in different labs. Alternatively, the difference might reflect a quantification caveat that arises when quantifying proteins that have accumulated under pre-steady-state conditions. Although the pSILAC experiment only quantified proteins synthesized after miRNA transfection, there still might have been a lag time before the impact of the transfected miRNA was fully exerted. Because pSILAC measured differences integrated over the time period of metabolic labeling, the maximal repression might have been underestimated if this lag time was substantial. The use of normal SILAC or direct-labeling methods for these and other pre-steady-state conditions would underestimate the maximal repression even more than would pSILAC. In contrast, ribosome profiling measures differences at one moment in time and thus is not subject to this quantification caveat. Consistent with this explanation, the protein changes tended to be at least as strong

as the ribosome changes in the miR-223 experiments (panel **c**), in which the cells that expressed miR-223 had been expressing this miRNA over many cell divisions, whereas those that had no miR-223 had never had this miRNA. Indeed, the changes appeared even greater in the previous experiment measuring protein changes compared to the current experiment measuring RPF changes (panel **c**). The simplest explanation for these differences is that the two biological samples differed, such that the miR-223 effects were greater when the experiment was done to collect the protein data compared to when the experiment was done to collect the RPF data (perhaps because in the first experiment samples were collected after eight days of differentiation, whereas in the second experiment samples were collected after only six days). Results from mRNA array data that had been collected on both biological samples, which showed greater mRNA differences in the first compared to the second experiment, supported this explanation (panel **d**). Therefore, the greater protein differences observed in panel **c** were due to greater effects in the first biological sample and were not evidence that the RPF changes failed to capture an aspect of miRNA regulation observed with the protein changes.

Taken together, the results of this figure indicate that changes observed by proteomics are no greater than those observed by ribosome profiling. Therefore, miRNAs do not change protein output or protein accumulation in a manner that might be missed by ribosome profiling, e.g., by co-translational degradation of nascent polypeptides, or by a reduction in the rate of translation initiation coupled with a correspondingly slower rate of elongation.

Chapter 3

A genome-wide study of miR-155-mediated repression in activated B cells

Huili Guo[1,2], Chanseok Shin[3], Daehyun Baek[4], Judit Villen[5], and David P. Bartel[1,2]

[1]Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[3]Department of Agricultural Biotechnology, Seoul National University, Seoul, 151-921, Republic of Korea
[4]School of Biological Sciences, Seoul National University, Seoul, 151-747, Republic of South Korea
[5]Department of Genome Sciences, University of Washington, Seattle WA, 98195, USA

H.G. performed mRNA-Seq, ribosome profiling and the associated data analyses. C.S. performed the mouse experiments. J.V. performed the mass spectrometry and associated computational analyses. D.B performed the computational analysis of targeting associated with the proteomics experiment. All authors contributed to the design of the study.

Abstract

MicroRNAs (miRNAs) are endogenous ~22-nucleotide RNAs that base pair to their target mRNAs to direct repression. Downregulation of protein expression output can be mediated by translational repression and/or mRNA destabilization. We have previously shown that lowered mRNA levels account for most of the decreased protein production for both ectopic and endogenous miRNA:target interactions. Here, we investigate another endogenous system, lipopolysaccharide-activated B cells, using parallel ribosome profiling, mRNA-Seq and proteomics approaches. While the majority of the repression mediated by miR-155, a miRNA that is strongly induced upon B cell activation, can be attributable to changes in mRNA levels, the fraction contributed by the translational repression component is ~29%, somewhat higher than the 11–16% that we had previously observed. We also find evidence of widespread usage of upstream open reading frames in these cells, translated from canonical and non-canonical initiation codons. These upstream open reading frames are also translationally repressed by miR-155.

Introduction

MicroRNAs (miRNAs) are ~22-nucleotide small RNAs that downregulate expression of their targets by base pairing to the target mRNAs, and in so doing, fine-tune gene expression (Bartel, 2009). With a few exceptions, targeting specificity in metazoans is mediated by nucleotides 2–7 of the 5′ region of the miRNA, known as the 'seed'. This seed region of the miRNA directs Argonaute, the core component in the effector complex that the miRNA is associated with, to the target mRNA to mediate repression (Bartel, 2009). Each conserved miRNA targets hundreds of distinct genes (Friedman et al., 2009), and thus, miRNAs have been shown to affect many cellular processes and impact a multitude of biological functions (Bartel, 2009; Bushati and Cohen, 2007).

Although it is clear that miRNAs play important gene-regulatory roles, how they do so has been more controversial. The first miRNA, *lin-4*, was shown to downregulate expression of its target *lin-14* by translational repression, without substantial change at the mRNA level (Olsen and Ambros, 1999; Wightman et al., 1993). However, it was later found that *lin-4* does downregulate *lin-14* mRNA levels (Bagga et al., 2005). Reduction in mRNA levels was also observed, by microarray, for transcripts with complementary sites to ectopically expressed miRNAs in HeLa cells (Lim et al., 2005). Conversely, the levels of mRNAs with predicted target sites increase when the cognate miRNA is depleted (Krutzfeldt et al., 2005). Subsequently, it was shown that during the maternal-to-zygotic transition in zebrafish embryos, miR-430 downregulated hundreds of maternal transcripts by directing their deadenylation (Giraldez et al., 2006). These data indicated that miRNA-mediated mRNA destabilization is a more widespread phenomenon than previously appreciated. At the same time, single-gene reporter studies indicated that different miRNA:target interactions can lead to different amounts of mRNA destabilization and/or translational repression (Eulalio et al., 2008). The question raised then was, at the genome-wide level, to what extent does translational repression and mRNA destabilization each contribute towards the overall repression?

To address the worry that targets that are predominantly translationally repressed might have been missed by microarray studies, high-throughput studies employing proteomics approaches were

carried out with simultaneous mRNA measurements by microarrays (Baek et al., 2008; Selbach et al., 2008). From such studies, it was found that changes in mRNA abundance could explain much of the change observed at the protein level (Baek et al., 2008). However, proteomics approaches can be biased towards detecting highly abundant proteins and lack the wide dynamic range of genomics approaches. Monitoring steady-state protein levels also means that potential differences in protein degradation rates would be concurrently reflected in the measurements, potentially confounding inferences made of protein synthesis. In an attempt to bypass the need to measure protein levels altogether, polysome profiling was used, together with a gradient-encoding method, followed by downstream microarray analysis, to investigate repression after ectopic expression of miR-124 in HEK293 cells (Hendrickson et al., 2009). This study found that target mRNA reduction could explain ~75% of the total amount of repression, though it still left open the question of whether endogenous miRNA:target interactions behave in the same manner.

We have previously used ribosome profiling, together with parallel mRNA-Seq measurements, to address this question. This approach was used to investigate three different miRNAs, in two different systems. We had transfected miR-155 and miR-1 into HeLa cells, which normally do not express these two miRNAs; we also investigated an endogenous system — *in vitro* differentiated neutrophils from *mir-223* knockout versus wild-type mice. In all three cases, we found that ≥84% of the overall repression mediated by miRNAs can be explained by a reduction in steady state mRNA levels (Guo et al., 2010). Although the neutrophils were a true endogenous system, we wondered if their short lifespan (Brinkmann and Zychlinsky, 2007) might make them an outlier in the range of mammalian cell types. To test the generality of our initial findings, we decided to investigate another endogenous system.

Mice deficient in the *bic/mir-155* gene are viable and fertile, but exhibit defects in B and T cell immunity (Rodriguez et al., 2007; Thai et al., 2007). The expression of miR-155 is strongly induced in activated B and T cells, as well as in activated macrophages and dendritic cells (O'Connell et al., 2007; Rodriguez et al., 2007; Thai et al., 2007), underscoring the importance of miR-155 in adaptive, as well as innate immunity.

We have performed mRNA-Seq, ribosome profiling and pulsed stable isotope labeling with amino acids in cell culture (pSILAC) in activated B cells from *mir-155* knockout and wild-type mice, and compared gene expression changes at the mRNA, translation, and protein levels. We found that the majority of the repression mediated by miR-155 can be explained by mRNA destabilization effects, although a larger component is attributable to translational repression than we had previously seen (Guo et al., 2010). Translational repression of upstream open reading frames (uORFs) was also observed.

Results

Translational status of activated B cells

As an alternative to the neutrophil system, we chose to investigate *mir-155* knockout versus wild-type mice. We stimulated splenic B cells with lipopolysaccharide (LPS), interleukin-4 (IL-4) and anti-CD40 for 48 hours and performed ribosome profiling and mRNA-Seq in parallel to look at the impact of miR-155 on the transcriptome. In our original study, we had made use of previously published proteomic data sets to filter out a set of target genes that are likely to be enriched for functional target sites, and based our analyses on these sets of genes that are validated by proteomic data (Guo et al., 2010). Although the use of proteomics-supported targets gave us more confidence that we were looking at true direct targets, this method can be improved if the protein measurements were to come from the exact same samples used in ribosome profiling and mRNA-Seq. As such, we also performed pSILAC on the same cells, with the pulse labeling proteins synthesized after LPS exposure (Figure S1A). Consistent with previous reports, miR-155 in these cells was strongly upregulated after LPS exposure (Figure S1B).

Upon activation, B cells eventually exit the cell cycle and undergo terminal differentiation (Shapiro-Shelef and Calame, 2005). Although the eventual fate of activated B cells is to differentiate into highly secretory plasma cells, the upregulation of protein synthesis is not linear (Gass et al., 2008; Goldfinger et al.; van Anken et al., 2003). In addition, different groups of functionally related proteins, such as protein folding chaperones and metabolic enzymes, are sequentially expressed in anticipation of the professional secretory role of the cells (van Anken et al., 2003). Massive immunoglobulin production

takes place only after an initial time lag. From polysome profiling, we found that unlike in actively-dividing cells, the polysome:monosome ratio was much lower in activated B cells (Figure 1A, compared to Figure S2A). There were clearly ribosomes present in the higher polysomal fractions, but these were found in much lower proportions, compared to the monosome fractions (Figure 1B). Plotting the ribosome footprint density profile for detected genes revealed that a substantial proportion of ribosome-protected mRNA fragments (RPFs) is located at the position characteristic of where the ribosome would be found when it contains the start codon in its P site (Figure 1C). Compared to the steady state density level, the density at this peak is ~40-fold higher. In our previous study, the analogous ratio for HeLa cells and neutrophils was ~5-fold. Because the density at this peak is likely to saturate (see Figure 1C legend), the large increase in ratio in the B cells suggests that ribosome loading is much lower in activated B cells, compared to actively-dividing neutrophils or HeLa cells. This corroborates the low polysome:monosome ratio observed in the polysome profile (Figure 1A). In addition, ~14% of the RPFs that map to the coding region of detected genes map to this initial peak (Table S1). This contrasts with the analogous ~2% seen in our previous study.

Because the monosomes make up a much larger proportion than 14% in the polysome profile (Figure 1A and 1B), this indicates that if most of the 80S ribosomes were associated with an RPF, some of these RPFs must have derived from regions outside the start codon. This would suggest that despite not being polysomal, a large portion of the monosomes might be engaged in productive translation. Intriguingly, when Raji cells, which are mature B cells from a human Burkitt's lymphoma cell line, were transduced to differentiate into plasma cells, the number of assembled 80S ribosomes increased, but the overall loading of mRNAs onto polysomes was described as "extremely low and therefore difficult to quantitate by ultracentrifugation" (Shaffer et al., 2004). Alternatively, it is possible that many of these 80S ribosomes are not loaded onto mRNAs and are thus 'empty' monosomes without associated RPFs. Either scenario would mean that the amount of translation in activated B cells is much lower than that in HeLa cells or neutrophils. To prevent the large number of RPFs mapping to this initial peak (Figure 1C)

from skewing our results, we only used RPFs that map to the coding region from codon 9 onwards for subsequent analyses.

Plotting the RPF expression values against expression values derived from counting mRNA-Seq tags indicates that a group of genes was translationally downregulated in these cells (Figure 1D). The same genes were not visibly downregulated translationally in the *in vitro* differentiated neutrophils (Figure S2B). This group of genes is enriched for ribosomal proteins, as well as factors involved in protein synthesis (Figure 1E). Messages encoding ribosomal proteins are known to be translationally regulated by a 5′-terminal oligopyrimidine tract (5′TOP), which is usually made up of a terminal cytosine moiety, followed by a run of pyrimidines (Meyuhas, 2000). Besides ribosomal protein genes, genes encoding human homologs of translation elongation factors and three subunits of the initiation factor eIF3 have also been validated as 5′TOP mRNAs (Iadevaia et al., 2008). Thus, of the 40 genes in this translationally downregulated group, 27 are likely to be 5′TOP mRNAs, including three elongation factors and *Eif3f*. A closer inspection of the 5′-untranslated regions (5′UTRs) of the remaining 13 genes indicated that all except three contain motifs that could serve as 5′TOP motifs (Figure 1E). Interestingly, 6 of these 13 genes encode proteins that have a role in translation, or proteins that are involved in protein folding, which could also impact translation (Figure 1E). Because 5′TOP mRNAs are typically downregulated when global protein synthesis is reduced (Meyuhas, 2000), this suggests that at 48 hours post-stimulation, the B cells are unlikely to be as translationally active as the actively-dividing neutrophils. Again, this is consistent with the observed polysome profile (Figure 1A). Interestingly, in dendritic cells stimulated with LPS, it has been observed that overall protein synthesis levels first increase, then decrease at a later time point. In particular, messages encoding ribosomal proteins were specifically disengaged from polysomes at a late stage of maturation (Ceppi et al., 2009). It may be that upon antigen stimulation, both B cells and dendritic cells employ similar mechanisms to regulate translation on their way to terminal differentiation.

mRNA destabilization is the predominant component of repression mediated by miR-155

We next examined the molecular consequences of miR-155 targeting, for genes with at least one canonical 7–8mer site to miR-155 in their 3′UTRs (Figure S3A). At both the mRNA-Seq level and the RPF level, the extent of derepression caused by deleting *mir-155* mirrored the established hierarchy of target site efficacies (Bartel, 2009)(Figure S3B). Using the parallel pSILAC data, we highlighted a subset of genes that were detected by proteomics (proteomics-detected targets). In addition, out of these proteomics-detected targets, we filtered out a set of genes that are likely to be enriched for functional target sites, by virtue of their robust response at the protein level (pSILAC $\log_2$ fold change ≥ 0.3, proteomics-supported), and focused our analyses on this set of targets. These proteomics-supported targets responded more robustly to the deletion of *mir-155* (Figure 2A), validating the use of this approach. Because changes at the ribosome profiling level would encompass changes at the underlying mRNA level, derepression at the level of steady state translational efficiency can only be derived after normalizing the RPF change by the mRNA-Seq change. The residual change in translational efficiency obtained after normalization can be thought of as a change in the number of ribosomes per mRNA. We observed a statistically significant increase in translational efficiency for the proteomics-supported targets in response to *mir-155* deletion (Figure 2B; $P < 10^{-5}$), indicating that miR-155 targeting leads to reduced ribosome density. The fraction of repression observed by ribosome profiling that could be attributable to reduced translational efficiency was ~29%, compared to the 11–16% that we had previously observed (Guo et al., 2010). Notably, the hierarchy of site efficacy was preserved even after the normalization (Figure S3C). This increased contribution from translational repression was not due to a few outlier genes being much more translationally repressed than others because correlation between RPF changes and mRNA-Seq changes was strong for genes with miR-155 target sites (Figure 2C; $R^2 = 0.53$), and this strong correlation was not observed in parallel analysis of genes without sites (Figure 2D; $R^2 = 0.30$).

Translation of upstream ORFs

A closer look at the density profile of individual genes highlighted the fact that many of the RPFs mapping to the 5′UTR appear to cluster in peaks over certain locations, much like the characteristic peaks formed by ribosomes stalled at the start codons of the coding region. But unlike the peaks at the start codons of the ORFs, many of these peaks were not centered over AUGs. An example is the gene *Rela*, with peaks centering over potential upstream ORFs (uORFs) initiated from an upstream AUG, as well as from non-canonical CUG and GUG (Figure 3A). This discrepancy becomes more obvious after building sequence logos of aligned 30-mer reads, which are most representative of *bona fide* ribosome footprints, as the typical length of a mammalian ribosome footprint is 30 nucleotides (nt). A 30-nt RPF from a ribosome that is stalled at the start codon, with the start codon in its P site, would be expected to contain the start codon at positions 13–15, counting from the first base of the RPF (Figure 3B). Indeed, the sequence logo generated from 30-nt RPFs mapping to the 5′-most 5% of detected ORFs tend to contain 'AUG' at positions 13–15 (Figure 3C). The increased likelihood of having an 'A' at position 10 and a 'G' at position 16 of a 30-nt RPF translates directly to the Kozak sequence, the presence of which enhances the efficiency of translation initiation (Kozak, 1986, 1987). The same motif can also be observed after aligning 30-nt RPFs mapping to detected ORFs in the neutrophil data set from our previous study (Figure S4A).

When an analogous sequence logo was built for RPFs mapping to 5′UTRs in the neutrophil data set, no motif was seen (Figure S4B). In contrast, 30-nt RPFs mapping to 5′UTRs in activated B cells tend to have a 'T' at position 14, followed by a 'G' at position 15 (Figure 3D), suggesting the usage of non-canonical start codons in 5′UTRs. Interestingly, no obvious Kozak sequence was observed. These 5′UTR RPFs are likely to have derived from fully assembled 80S ribosomes, rather than scanning 40S small subunits. 40S subunits tend to protect longer fragments of mRNA than 80S ribosomes (Kozak and Shatkin, 1977; Lazarowitz and Robertson, 1977), likely owing to associated initiation factors during the scanning process. Such longer fragments would have been discarded at the size selection step during the

preparation of the RPF libraries. This would suggest that like in yeast (Ingolia et al., 2009), usage of uORFs with non-canonical start codons might be more prevalent in mammalian cells than previously appreciated. This is also consistent with a similar observation made recently in mouse embryonic stem cells (Nicholas Ingolia, personal communication). The lack of any observable motif from RPFs mapping to the 5′UTRs in the neutrophil data set does not, however, imply that uORF translation is less common in neutrophils. This is because the lack of a strong initial peak over start codons in the neutrophil data set precluded the possibility of seeing a motif by sequence logo analysis. Conversely, because we expect all used start codons in activated B cells to be 'marked' by the strong initial peak seen in the main ORFs (Figure 1C), this characteristic can be used to identify and annotate *bona fide* start codons in the 5′UTRs.

The uORFs annotated by this approach appear to be capable of initiating productive translation. Productive translation is characterized by a 3-nt periodicity that reflects the codon-by-codon movement of translating ribosomes, and this is indeed observed after combining RPFs from annotated AUG-uORFs into a composite uORF (Figure S5A). Successful 80S formation is also reflected by a peak at position +3, relative to the first base (numbered '0') of the uORF. These characteristic patterns exhibited by AUG-uORFs enabled us to verify the other candidate uORFs that initiated from non-canonical start codons (Figure S5B). From these density profiles, it is clear that AUG-uORFs are the most highly translated uORFs. This is not inconsistent with the lack of emergence of an AUG motif in the 5′UTR sequence logo (Figure 3D), because while a sequence logo gives a summary of the most commonly found bases in each position, the density profiles give the read density at each position, taking into account the frequency of uORFs. For example, CUG is the top codon present at positions 13–15 of the sequence logo (Table S2), but there are many more CUG-uORFs than AUG-uORFs. Thus, after normalizing for the frequency of uORF occurrence, AUG-uORFs are the most highly expressed, followed by CUG-uORFs (Figure S5B). It was previously reported that during antigen presentation, CUG start codons can be decoded as leucines, in addition to methionines (Schwab et al., 2004). However, if this occurs in the activated B cells, it is extremely uncommon because after ranking the initiation codons based on translational efficiency, as

judged by the density profiles (Figure S5B), we obtained a hierarchy of AUG >> CUG > GUG/UUG/ACG > AUA > AUU > AUC >> AGG/AAG, which is near-identical to that reported in a study that used methionyl initiator tRNA (Met-tRNA$_i$) to probe the kinetics and thermodynamics of codon:anticodon base pairing (Kolitz et al., 2009). This suggests that the vast majority of uORF initiation codons in activated B cells are likely to be decoded by Met-tRNA$_i$.

Translational repression of upstream ORFs mediated by miR-155

With the annotated uORFs, we can then proceed to ask if the translation of uORFs of genes with miR-155 target sites was also repressed. An approach similar to that used to interrogate repression in the main coding region was applied here, for genes with uORFs in their 5′UTRs, and whose uORFs have a sufficient number of mapped RPFs to pass our threshold for quantification. An aggregate RPF change was derived for each gene based on the sum of reads that mapped to all the uORFs of a given gene. Doing the equivalent for mRNA-Seq tags would yield too few tags to accurately quantify the aggregate uORF gene expression. Hence, we derived the mRNA-Seq change based on mRNA-Seq tags that map to the entire 5′UTR of a given gene. This method gave statistically significant derepression at the mRNA and RPF levels (Figure 4A; $P < 0.01$ and $P < 10^{-8}$, respectively). The former is not unexpected because changes in mRNA expression should be similar whether quantified by tags in the coding region, or tags in the 5′UTR. After normalizing the aggregate-uORF RPF change by the 5′UTR mRNA change, we observed statistically significant derepression at the level of translational efficiency in the uORFs (Figure 4B; $P < 0.01$).

Another way to visualize changes in translational efficiency in the 5′UTR is to plot gene expression changes across a composite 5′UTR. We had previously applied this approach to the coding region of detected genes to probe for evidence of ribosome drop-off (Guo et al., 2010). Applying the same approach to detected 5′UTRs in the current data set, we observed derepression of RPFs above that attributable to changes at the mRNA level, further supporting the observation that miR-155 could also reduce ribosome density in the 5′UTR (Figure 4C). As >93% of all 5′UTR RPFs map to the annotated

uORFs (Table S1), these results suggest that the decrease in ribosome density in the 5′UTR occurs primarily in the uORFs.

Discussion

We have shown that in activated B cells, miR-155 primarily exerts its effects at the mRNA level, consistent with what was previously observed in HeLa cell transfections, and *in vitro* differentiated neutrophils (Guo et al., 2010). Although the majority of the impact is exerted at the mRNA level, the fraction that can be attributed to translational repression (Figure 2; ~29%) was somewhat higher than the 11–16% that was previously observed. In our previous study, we had brought up the possibility that this component could be an artifact of using total RNA, instead of cytoplasmic RNA, for mRNA-Seq, which could lead to apparent changes in translational efficiency — even if there was actually no translational repression — due to nuclear sequestration of mRNAs. However, we have since performed mRNA-Seq using cytoplasmic RNA (in addition to total RNA) in U2-OS cell transfections. For both miR-155 and miR-1 transfections, the component from translational efficiency changes remained whether total RNA or cytoplasmic RNA was used for mRNA-Seq, indicating that this component is unlikely to be the result of nuclear sequestration, and is probably *bona fide* translational repression (data not shown).

As reported previously, this reduction in ribosome density is likely to be due to inhibition of translation initiation. We have previously shown that cotranslational degradation, or a mixture of initiation and elongation blocks, which are both mechanisms that could reduce protein output without changing ribosome density, are unlikely mechanisms of repression. This was because they would predict that proteomics-measured changes would exceed those measured by ribosome profiling, and that was not observed in the previous study. In the current data set, in which we have proteomic data for the exact same biological samples, we can say so with even higher confidence (Figure S6). The translational repression observed in uORFs is also consistent with the inhibition of translation initiation, such as by reducing ribosome loading. Upstream ORF translation has previously been observed to correlate with reduced expression from the main coding region (Calvo et al., 2009), thus it would be interesting to

investigate potential regulatory interplay between this effect and miRNA-mediated repression (also see Chapter 4).

This study demonstrates that the molecular consequences effected by miRNAs can be slightly different in different cell types. miR-155 was used in the HeLa cell transfections in the previous study, thus the difference in contributions from the mRNA destabilization and translational repression components cannot be attributed to a difference in miRNA identity. It is interesting to note that the cell types used in the previous study were cells that were actively dividing, in contrast to the LPS-stimulated B cells, which were not. Exploiting such differences in an attempt to investigate the disparity in miRNA repression between HeLa cells/neutrophils and activated B cells would be the basis for future directions (see Chapter 4).

Methods

B cell isolation and activation

All animal experiments were approved by the MIT Committee on Animal Care. Spleen was obtained from four 3-month-old wild-type male mice and from four 3-month-old *mir-155*$^{-/-}$ mice, and B cells were harvested as follows. Spleen from the four mice of each genotype was pooled, and single cells were isolated after ammonium chloride treatment and incubation with anti-Mouse CD45R/B220 magnetic particles (BD Biosciences). Isolated B cells were then resuspended in RPMI media. B cells from wild-type mice were activated in RPMI media supplemented with L-Lysine:2HCl (4,4,5,5-D4; Cambridge Isotope Laboratories, Inc.) and containing lipopolysaccharide (LPS), IL-4, and anti-CD40 antibody. B cells from *mir-155*$^{-/-}$ mice were activated in RPMI media supplemented with L-Lysine:2HCl (U-13C6, U-15N2; Cambridge Isotope Laboratories, Inc.) and containing LPS, IL-4, and anti-CD40 antibody. Activation was done in media supplemented with lysines of different molecular weights for the pSILAC experiment. At 48 hours after LPS exposure, cycloheximide (100 μg/ml) was added to arrest translation. After incubating for 10 min at 37 °C, cells were harvested in ice-cold PBS supplemented with

cycloheximide and split into three portions for mRNA profiling, ribosome profiling and proteomics. Equal numbers of cells from each genotype were mixed for the proteomics analysis.

Ribosome footprinting and RPF purification

Ribosome footprinting and RPF purification was performed as described (Guo et al., 2010), with the following modifications. After pooling monosome fractions, proteinase K digest was carried out without applying the fractions through a membrane filter to remove non-covalently associated ribosomal RNA (rRNA). Instead, an additional subtractive hybridization step was used to remove rRNA contaminants during preparation of the RPF libraries.

mRNA fragmentation and mRNA-Seq library preparation

mRNA fragmentation and library preparation for mRNA-Seq was performed as described (Guo et al., 2010), except that mRNA fragments were size-selected from 33–53 nt.

RPF library preparation

RPF libraries were prepared as described (Guo et al., 2010) but with the addition of a subtractive hybridization step, to remove rRNA contaminants, between the 3′-ligation and 5′-phosphorylation steps. After ligation of the 3′ adaptor, the ligation reaction was phenol/chloroform-extracted and ethanol precipitated. The RNA from the 3′-ligation reaction was then incubated in 2 x SSC with biotinylated oligonucleotides that are reverse complement sequences of abundant rRNA contaminants observed in preliminary sequencing experiments. The mixture was heated at 70 °C for 5 min, and slow-cooled to allow annealing. Biotinylated oligonucleotides were then recovered using MyOne streptavidin C1 DynaBeads (Invitrogen) according to manufacturer's instructions. The RNA remaining was then precipitated and gel-purified before the 5′-phosphorylation step.

Oligonucleotide sequences

All sequences listed are DNA oligonucleotides (listed in 5′ to 3′ orientation) biotinylated at the 3′ end, with the biotin attached to the 3′ terminus by a C6 spacer.

HG_asDNA_seed1: CGAGGTTATCTAGAGTCACCA-(biotin)

HG_asDNA_seed2: TCCTAGCTGCGGTATCCAGGCG-(biotin)

HG_asDNA_seed3: TAGAATTACCACAGTTATC-(biotin)

HG_asDNA_seed4: TCAGAAGGACTTGGGCCCCCC-(biotin)

HG_asDNA_seed5: TTGGCGCCAGAAGCGAGAGCC-(biotin)

HG_asDNA_seed6: ATAAATGCACGCATCCCCC-(biotin)

HG_asDNA_seed7: CTATCCGGGGCCAACCGAG-(biotin)

Pulsed SILAC

Mass spectrometry and peptide quantification analyses were performed as described in Baek et al., 2008, except that a different in-house software (CoreQUANT) and second isotopes were used for quantification.

Sequence analyses

Illumina sequencing reads were mapped using the Bowtie mapping program (Langmead et al., 2009), with the same iterative mapping strategy previously described (Guo et al., 2010), but with the following modifications. The first 29 nucleotides of each read were used as the 'seed' region, instead of the first 25. After mapping reads with one seed mismatch, reads that failed to map were fed back into Bowtie again, to allow mapping of reads with two seed mismatches. A mouse reference transcript database was compiled as previously described, starting from a refFlat file generated on 26 December, 2010, which was downloaded from the UCSC Genome Browser (http://genome.ucsc.edu).

Quantification of gene expression

Gene expression was quantified as described, except that only RPFs and mRNA-Seq tags that map to the coding region from codon 9 onwards were used, to avoid any influence from the strong initial peak at the start codons (Figure 1C). Hence, in calculating the rpkM values, the feature length used was the length of the coding region, excluding the first 24 nucleotides. Quantification of fold changes in uORF expression was performed as described in Figure 4A. For each gene, the aggregate-uORF fold change in RPFs was calculated based on RPFs that mapped to the annotated uORFs, and the fold change in mRNA-Seq tags was calculated based on tags mapping to the entire 5′UTR.

Annotation of upstream open reading frames

First, 5′UTRs from the reference transcript database were scanned for all possible open reading frames that initiate from canonical AUG codons, as well as non-canonical start codons that were one base-change away from AUG. Then, these candidate uORFs were filtered for usage in the activated B cells. To obtain the set of uORFs that were actually translated, we made use of the strong initial peak at start codons (Figure 1C). A ribosome paused over a start codon, with the start codon in its P site, would generate an RPF whose 5′ terminus would map to position +3 relative to the first base (numbered '0') of the candidate uORF, after shifting positions +15 nt to reflect the position of the first nucleotide in the ribosome A site. Thus, if a candidate uORF has at least 5 reads mapping to this position, it is annotated as a translated uORF. In addition, because 30-nt RPFs are most representative of *bona fide* ribosome footprints, a uORF with fewer than 5 reads mapping to position +3 can still be annotated as a translated uORF if it has at least three 30-nt RPFs mapping to this position. These two criteria were used to annotate translated uORFs. Subsequently, AGG- and AAG-initiated uORFs were excluded from the set of uORF annotations because the translation initiation efficiencies of these two codons were much lower relative to the other codons (Figure S5B).

Figure legends

Figure 1. Translational status of activated B cells.

(A) Polysome profile of wild-type B cells that have been stimulated with LPS.

(B) Agarose gel of RNA extracted from corresponding fractions in the polysome profile in panel **A**.

(C) Density of RPFs and mRNA-Seq tags near the ends of ORFs in activated B cells, combining data from all quantified genes. Plotted are the 5′ termini of RPFs and mRNA-Seq tags, as reads per million reads mapping to genes (rpM). For RPF density, positions are shifted +15 nucleotides to reflect the position of the first nucleotide in the ribosome A site. Composite data are shown for ≥600-nucleotide ORFs that passed our threshold for quantification (≥100 RPFs and ≥100 mRNA-Seq tags).

The peak at the start codon is observed due to the use of cycloheximide, which is an elongation inhibitor that does not block translation initation. As such, even when elongation is completely blocked, 40S small subunits can still be loaded from the 5′ end of the mRNA. If these subunits make it to the start codon, each would be joined by a 60S large subunit to form a fully assembled 80S ribosome that would then be blocked by cycloheximide as it attempts to elongate. Thus, the density at this peak could saturate during the time period in which the cells are incubated with cycloheximide. Because this peak is expected to saturate, this would suggest that the different rpM values for this initial peak between the B cells and the neutrophils (see Chapter 2) actually describe the same (or largely similar) population of mRNAs. In turn, this suggests that the steady state level of RPF density per gene in the B cells (~40 fold difference between initial peak density and steady state density) is much lower than that in the neutrophils (~5 fold difference).

(D) Correspondence between gene expression, as quantified by ribosome profiling and mRNA-Seq, in the same library. Expression values are in terms of reads per million reads mapped to genes per kilobase coding exon model (rpkM). The R^2 derived from Pearson's correlation is indicated. Because poly(A)-selected mRNA was used for mRNA-Seq, histone genes (green circles), which are not polyadenylated, have much higher RPF expression values compared to their quantified mRNA expression. A group of genes that are translationally downregulated is highlighted (yellow circles). This group is likely to be

translationally downregulated because their lowered RPF:mRNA-Seq expression ratios form an outlier group, when compared to corresponding ratios from the same genes in neutrophils (Figure S2B).

(E) Distribution of genes that are translationally downregulated in B cells. Genes that were translationally downregulated in activated B cells (panel **D**) are divided according to whether they encode likely 5′TOP mRNAs. Genes within each category is further divided according to their biological function. Examples of 5′UTRs from each category are shown. For each example, the first ten bases starting from the potential 5′TOP motif is shown. If the first 'C' does not correspond to the annotated transcription start site, it is indicated by preceding bases in smaller case (e.g. *Eef1a1*). This could be due to misannotation of the 5′UTR, or the possibility that the first 'C' of the potential 5′TOP motif is genuinely not at the 5′ terminus of the mRNA.

Figure 2. miR-155 downregulated gene expression mostly through mRNA destabilization.

(A) Cumulative distributions of mRNA-Seq changes (left) and RPF changes (right) after deleting *mir-155*. Plotted are distributions for the genes with ≥1 miR-155 3′UTR site (blue), the subset of these genes detected in the pSILAC experiment (proteomics-detected, red), the subset of the proteomics-detected genes with proteins responding with $\log_2$-fold change ≥ 0.3 (proteomics-supported, green), and the control genes, which lacked miR-155 sites throughout their mRNAs (no site, black). The number of genes in each category is indicated in parentheses.

(B) Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after deleting *mir-155*. For each gene, the translational efficiency change was calculated by normalizing the RPF change by the mRNA-Seq change. For each distribution, the mean $\log_2$-fold change (±standard error) is shown (inset).

(C) Correspondence between ribosome (RPF) and mRNA (mRNA-Seq) changes after deleting *mir-155*, plotting data for the 683 quantified genes with at least one miR-155 3′UTR site (blue circles). Proteomics-detected targets and proteomics-supported targets are highlighted (pink diamonds and green crosses, respectively). Expected standard deviations (error bars) were calculated based on the number of

reads obtained per gene and assuming random counting statistics. The R^2 derived from Pearson's correlation of all data are indicated.

(D) Correspondence between ribosome and mRNA changes after deleting *mir-155*, plotting data for 683 genes randomly selected from the 3,914 quantified genes lacking a miR-155 site anywhere in the mRNA. Otherwise, as in panel **C**.

Figure 3. Upstream ORF translation occurs from canonical and non-canonical start codons in activated B cells.

(A) Read density from the 5′UTR of the *Rela* gene. Normalized read density (mRNA-Seq, blue and RPF, green) for *Rela*, is shown for the wild-type and *mir-155* knockout data sets. Note the five-fold difference in the range of the y-axes between the density plots for the RPF and mRNA-Seq data. Annotation tracks for potential uORFs are also shown. Characteristic peaks, as seen over start codons in the main ORFs (Figure 1C) are seen directly over potential upstream start codons. Note that only genome-matching reads were used to generate this coverage plot. Because the start codon of *Rela* is very close to a splice junction, the short length on the 5′ side of the junction precluded the mapping of genome-matching reads to this location by the mapping algorithm. This explains why a similar peak is not seen over the start codon of the main *Rela* ORF in this coverage plot.

(B) Schematic diagram of a ribosome with an associated 30-nucleotide RPF. The E, P and A sites of a ribosome, with the start codon in the P site, are shown. The expected dimensions illustrate why ribosomes stalled at start codons form characteristic peaks, with the start codons located in the middle of the peaks, as seen in panel **A**.

(C) Sequence logo generated from 30-nucleotide RPFs mapping to the ORFs. The sequence logo is built from aligned 30-nt RPFs that mapped to the 5′-most 5% of the coding regions of detected genes (pink bar). The expected location of the P site of the ribosome is indicated (green bar).

(D) Sequence logo generated from 30-nucleotide RPFs mapping to the 5′UTRs. The sequence logo is built from aligned 30-nt RPFs that mapped to 5′UTRs of detected genes (purple bar). The expected location of the P site of the ribosome is indicated (green bar).

Figure 4. Upstream ORFs are translationally repressed by miR-155.

(A) Cumulative distributions of mRNA-Seq changes (left) and RPF changes (right) in the uORFs after deleting *mir-155*. Only genes with uORFs that passed our threshold for quantification (for each gene, ≥10 RPFs from all its uORFs and ≥10 mRNA-Seq tags in its 5′UTR) are included in this plot. In addition, genes with uORFs that overlap the main coding region were excluded. Changes in mRNA levels were calculated from all mRNA-Seq tags that map to 5′UTRs. RPF changes were calculated using only RPFs that map to annotated uORFs. If a gene has more than one uORF, RPFs from its uORFs were summed before calculating a fold change. Plotted are distributions for the genes with ≥1 miR-155 3′UTR site (blue), the subset of these genes detected in the pSILAC experiment (proteomics-detected, red), the subset of the proteomics-detected genes with proteins responding with $\log_2$-fold change ≥ 0.3 (proteomics-supported, green), and the control genes, which lacked miR-155 sites throughout their mRNAs (no site, black). The number of genes in each category is indicated in parentheses.

(B) Cumulative distributions of translational efficiency changes in the uORFs for the polyadenylated mRNA that remained after deleting *mir-155*. For each gene, the translational efficiency change was calculated by normalizing the aggregate-uORF RPF change by the mRNA-Seq change. For each distribution, the mean $\log_2$-fold change (±standard error) is shown (inset).

(C) Ribosome and mRNA changes along the percentage length of the 5′UTRs after deleting *mir-155*. The 5′UTR of each quantified gene was divided into ten bins based on 5′UTR length. Fold changes in RPFs and mRNA-Seq tags mapping to each bin were then plotted. To avoid any influence from ribosomes stalled over the start codon of the main coding region, reads mapping to the six nucleotides immediately upstream of the ORF were excluded. Otherwise, all reads that mapped to the 5′UTRs were

used. Changes in RPFs and mRNA-Seq tags for mRNAs with no site (grey and black, respectively) and for proteomics-supported targets (light and dark green, respectively) are shown.

Acknowledgements

We thank N.T. Ingolia for sharing unpublished data and the ribosomal RNA subtractive hybridization protocol. We also thank the Whitehead Institute's Genome Technology Core for sequencing.

References

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature *455*, 64-71.

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A.E. (2005). Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. Cell *122*, 553-563.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Brinkmann, V., and Zychlinsky, A. (2007). Beneficial suicide: why neutrophils die to make NETs. Nat Rev Microbiol *5*, 577-582.

Bushati, N., and Cohen, S.M. (2007). microRNA functions. Annu Rev Cell Dev Biol *23*, 175-205.

Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci U S A *106*, 7507-7512.

Ceppi, M., Clavarino, G., Gatti, E., Schmidt, E.K., de Gassart, A., Blankenship, D., Ogola, G., Banchereau, J., Chaussabel, D., and Pierre, P. (2009). Ribosomal protein mRNAs are translationally-regulated during human dendritic cells activation by LPS. Immunome Res *5*, 5.

Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008). GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. Nat Struct Mol Biol *15*, 346-353.

Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res *19*, 92-105.

Gass, J.N., Jiang, H.Y., Wek, R.C., and Brewer, J.W. (2008). The unfolded protein response of B-lymphocytes: PERK-independent development of antibody-secreting cells. Mol Immunol *45*, 1035-1043.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75-79.

Goldfinger, M., Shmuel, M., Benhamron, S., and Tirosh, B. Protein synthesis in plasma cells is regulated by crosstalk between endoplasmic reticulum stress and mTOR signaling. Eur J Immunol *41*, 491-502.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835-840.

Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. PLoS Biol *7*, e1000238.

Iadevaia, V., Caldarola, S., Tino, E., Amaldi, F., and Loreni, F. (2008). All translation elongation factors and the e, f, and h subunits of translation initiation factor 3 are encoded by 5'-terminal oligopyrimidine (TOP) mRNAs. RNA *14*, 1730-1736.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.
Kolitz, S.E., Takacs, J.E., and Lorsch, J.R. (2009). Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. RNA *15*, 138-152.
Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell *44*, 283-292.
Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res *15*, 8125-8148.
Kozak, M., and Shatkin, A.J. (1977). Sequences of two 5'-terminal ribosome-protected fragments from reovirus messenger RNAs. J Mol Biol *112*, 75-96.
Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M., and Stoffel, M. (2005). Silencing of microRNAs in vivo with 'antagomirs'. Nature *438*, 685-689.
Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.
Lazarowitz, S.G., and Robertson, H.D. (1977). Initiator regions from the small size class of reovirus messenger RNA protected by rabbit reticulocyte ribosomes. J Biol Chem *252*, 7842-7849.
Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769-773.
Meyuhas, O. (2000). Synthesis of the translational apparatus is regulated at the translational level. Eur J Biochem *267*, 6321-6330.
O'Connell, R.M., Taganov, K.D., Boldin, M.P., Cheng, G., and Baltimore, D. (2007). MicroRNA-155 is induced during the macrophage inflammatory response. Proc Natl Acad Sci U S A *104*, 1604-1609.
Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Dev Biol *216*, 671-680.
Rodriguez, A., Vigorito, E., Clare, S., Warren, M.V., Couttet, P., Soond, D.R., van Dongen, S., Grocock, R.J., Das, P.P., Miska, E.A., *et al.* (2007). Requirement of bic/microRNA-155 for normal immune function. Science *316*, 608-611.
Schwab, S.R., Shugart, J.A., Horng, T., Malarkannan, S., and Shastri, N. (2004). Unanticipated antigens: translation initiation at CUG with leucine. PLoS Biol *2*, e366.
Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58-63.
Shaffer, A.L., Shapiro-Shelef, M., Iwakoshi, N.N., Lee, A.H., Qian, S.B., Zhao, H., Yu, X., Yang, L., Tan, B.K., Rosenwald, A., *et al.* (2004). XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. Immunity *21*, 81-93.
Shapiro-Shelef, M., and Calame, K. (2005). Regulation of plasma-cell development. Nat Rev Immunol *5*, 230-242.
Thai, T.H., Calado, D.P., Casola, S., Ansel, K.M., Xiao, C., Xue, Y., Murphy, A., Frendewey, D., Valenzuela, D., Kutok, J.L., *et al.* (2007). Regulation of the germinal center response by microRNA-155. Science *316*, 604-608.
van Anken, E., Romijn, E.P., Maggioni, C., Mezghrani, A., Sitia, R., Braakman, I., and Heck, A.J. (2003). Sequential waves of functionally related proteins are expressed when B cells prepare for antibody secretion. Immunity *18*, 243-253.
Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-862.

Figure 1

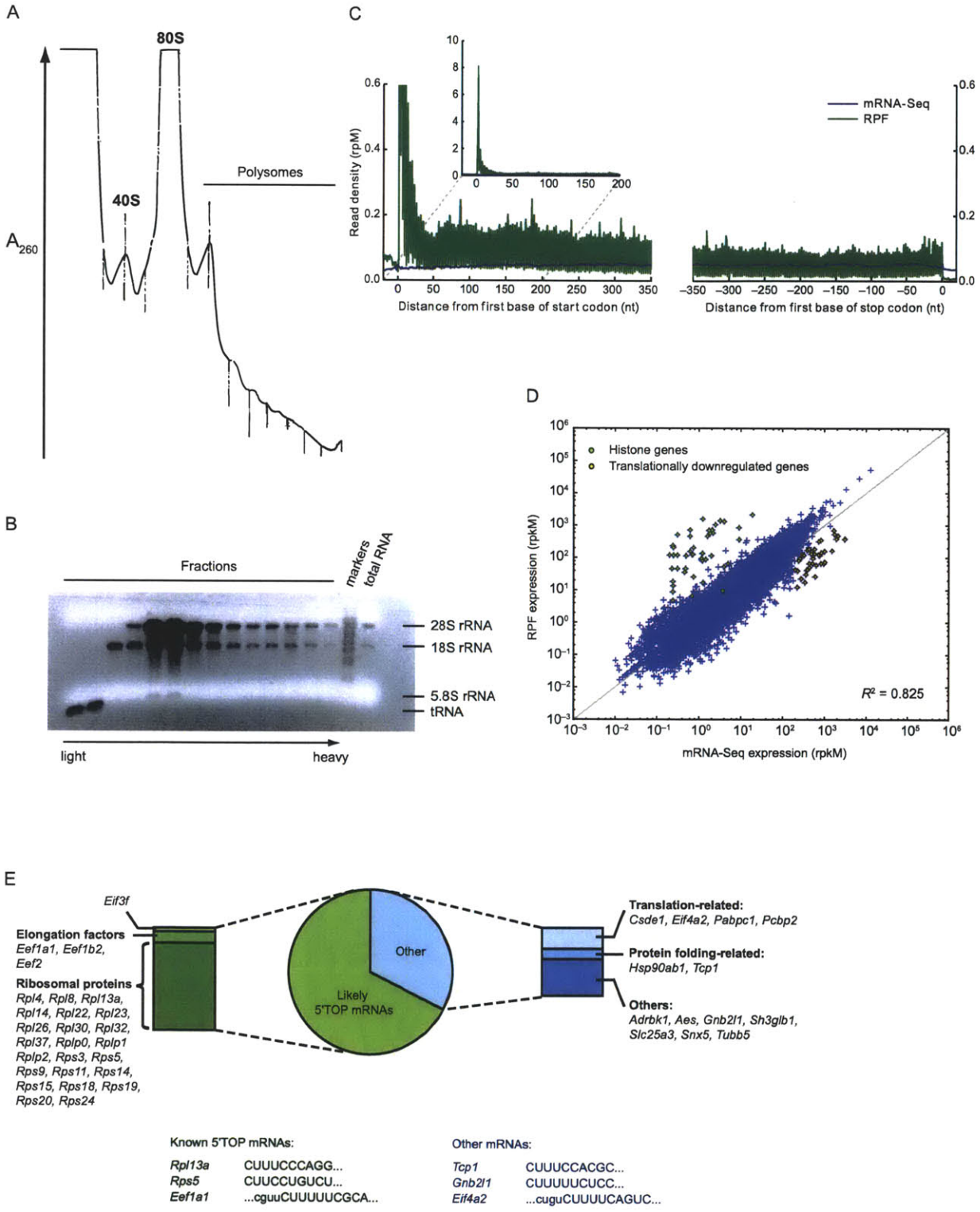

A
80S
40S
Polysomes
A260
B
Fractions
markers
total RNA
28S rRNA
18S rRNA
5.8S rRNA
tRNA
light
heavy
C
Read density (rpM)
Distance from first base of start codon (nt)
mRNA-Seq
RPF
Distance from first base of stop codon (nt)
D
Histone genes
Translationally downregulated genes
RPF expression (rpkM)
mRNA-Seq expression (rpkM)
R² = 0.825
E
Eif3f
Elongation factors
Eef1a1, Eef1b2, Eef2
Ribosomal proteins
Rpl4, Rpl8, Rpl13a, Rpl14, Rpl22, Rpl23, Rpl26, Rpl30, Rpl32, Rpl37, Rplp0, Rplp1 Rplp2, Rps3, Rps5, Rps9, Rps11, Rps14, Rps15, Rps18, Rps19, Rps20, Rps24
Other
Likely 5'TOP mRNAs
Translation-related:
Csde1, Eif4a2, Pabpc1, Pcbp2
Protein folding-related:
Hsp90ab1, Tcp1
Others:
Adrbk1, Aes, Gnb2l1, Sh3glb1, Slc25a3, Snx5, Tubb5
Known 5'TOP mRNAs:
Rpl13a CUUUCCCAGG...
Rps5 CUUCCUGUCU...
Eef1a1 ...cguuCUUUUUCGCA...
Other mRNAs:
Tcp1 CUUUCCACGC...
Gnb2l1 CUUUUUCUCC...
Eif4a2 ...cuguCUUUUCAGUC...

Figure 2

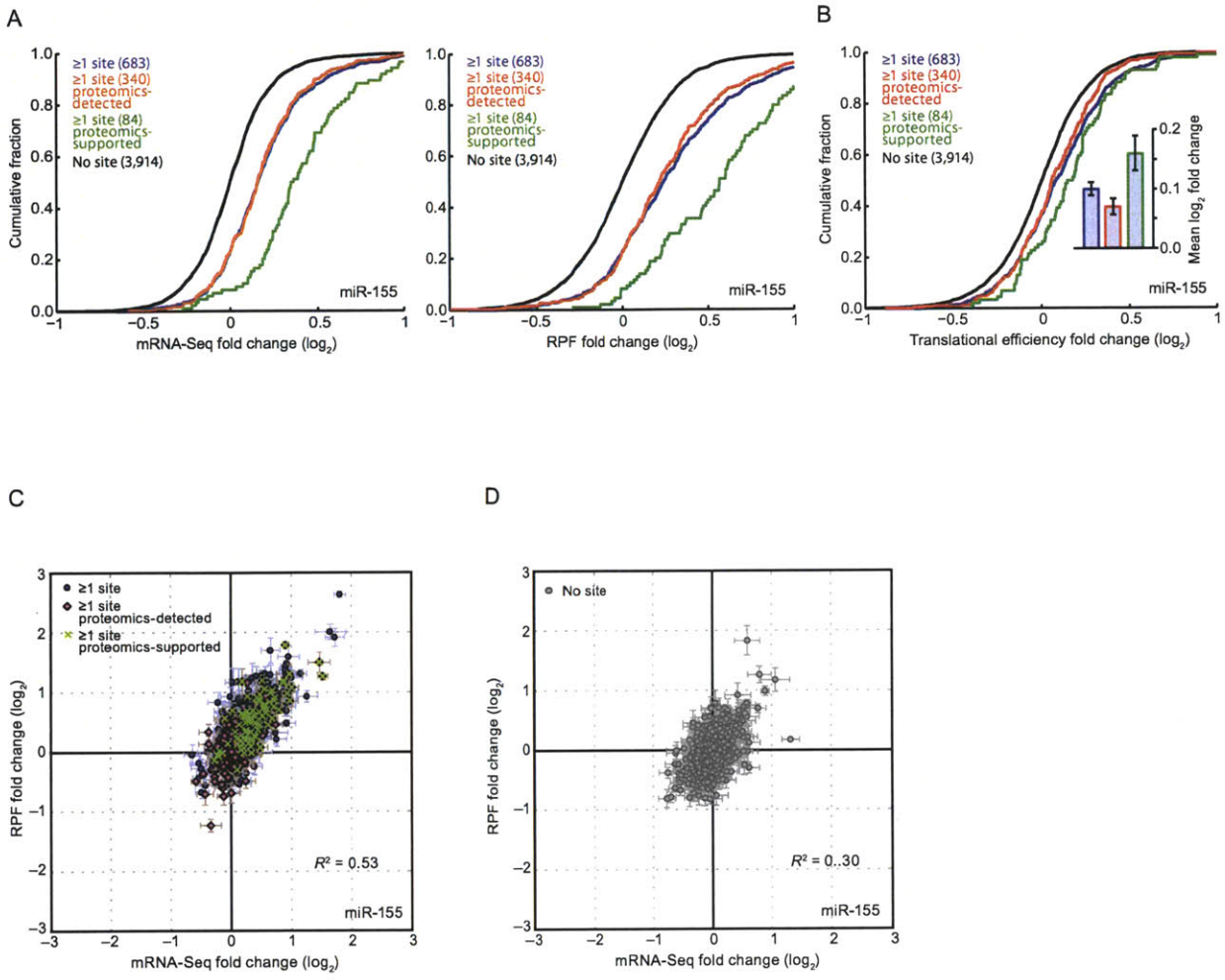

A
B
C
D
≥1 site (683)
≥1 site (340) proteomics-detected
≥1 site (84) proteomics-supported
No site (3,914)
Cumulative fraction
mRNA-Seq fold change (log2)
RPF fold change (log2)
Translational efficiency fold change (log2)
Mean log2 fold change
miR-155
≥1 site
≥1 site proteomics-detected
≥1 site proteomics-supported
No site
R² = 0.53
R² = 0..30

Figure 3

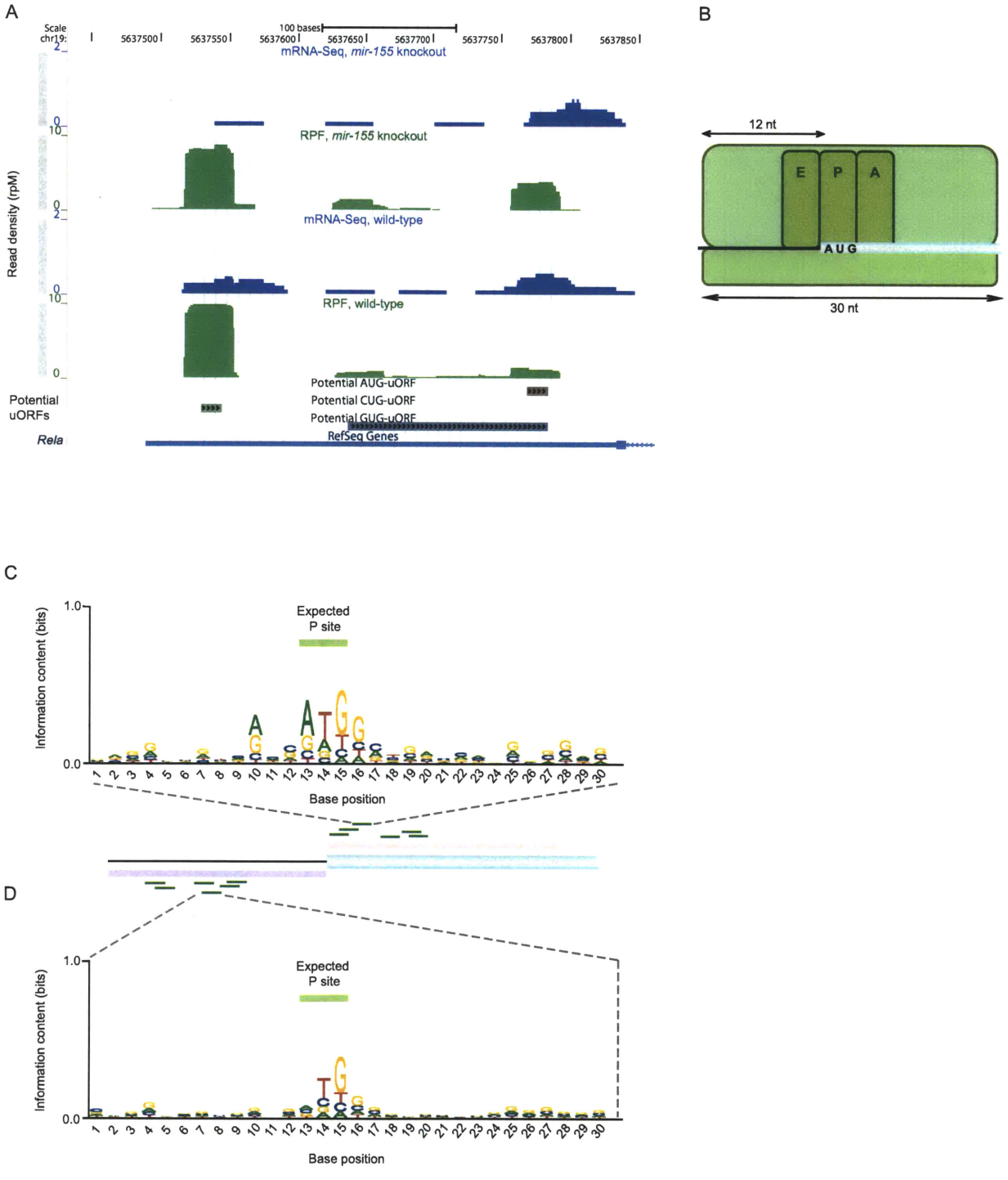
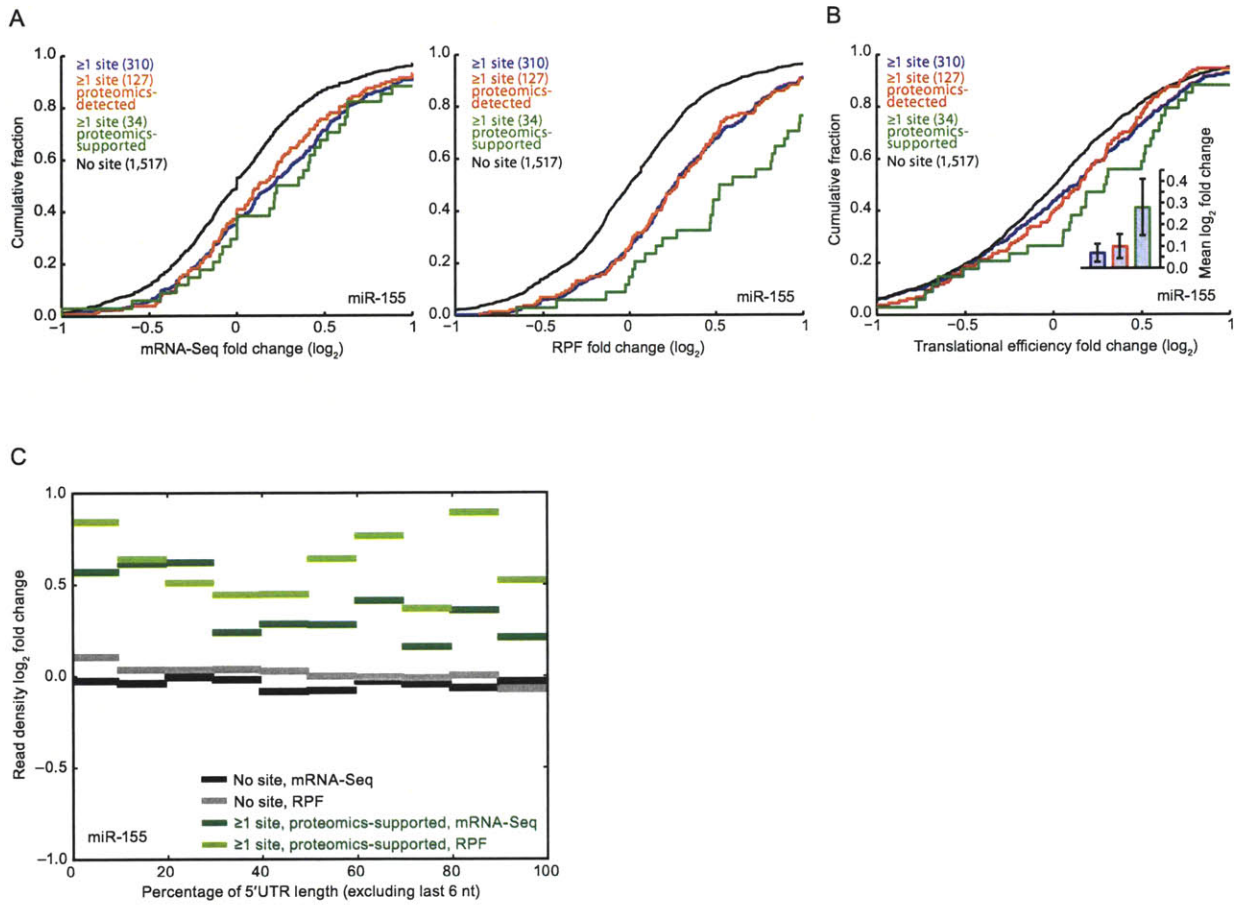

A
Scale
chr19:
100 bases
5637500
5637550
5637600
5637650
5637700
5637750
5637800
5637850
mRNA-Seq, mir-155 knockout
RPF, mir-155 knockout
mRNA-Seq, wild-type
RPF, wild-type
Read density (rpM)
Potential AUG-uORF
Potential CUG-uORF
Potential GUG-uORF
Potential uORFs
RefSeq Genes
Rela
B
12 nt
E
P
A
AUG
30 nt
C
Expected P site
Information content (bits)
Base position
D
Expected P site
Information content (bits)
Base position

Figure 4


A
B
C
≥1 site (310)
≥1 site (127) proteomics-detected
≥1 site (34) proteomics-supported
No site (1,517)
Cumulative fraction
mRNA-Seq fold change (log2)
RPF fold change (log2)
Translational efficiency fold change (log2)
Mean log2 fold change
miR-155
Read density log2 fold change
Percentage of 5′UTR length (excluding last 6 nt)
No site, mRNA-Seq
No site, RPF
≥1 site, proteomics-supported, mRNA-Seq
≥1 site, proteomics-supported, RPF

Supplemental materials

Supplemental figure legends

Figure S1. Activation of splenic B cells to induce miR-155 expression.

(A) Schematic of B cell activation and analysis.

(B) Expression of miR-155 before and after LPS activation. miR-155 expression was detected by Northern blotting. As a loading control, the blot was reprobed for U6 small nuclear RNA. The relative expression levels, normalized using the loading control, are shown below the blot.

Figure S2. Translational status of actively-dividing cells.

(A) Polysome profile of HeLa cells.

(B) Correspondence between gene expression, as quantified by ribosome profiling and mRNA-Seq, in the same neutrophil library. Expression values are in terms of reads per million reads mapped to genes per kilobase coding exon model (rpkM). The R^2 derived from Pearson's correlation is indicated. Because poly(A)-selected mRNA was used for mRNA-Seq, histone genes (green circles), which are not polyadenylated, have much higher RPF expression values compared to their quantified mRNA expression. A group of genes that were translationally downregulated in the activated B cells (Figure 1D) is highlighted (yellow circles).

Figure S3. Analyses of miR-155-dependent changes.

(A) Different types of seed-matched sites (Bartel, 2009).

(B) Cumulative distributions of mRNA-Seq changes (left) and RPF changes (right) after deleting *mir-155*. Plotted are distributions for genes with either a single 3'UTR site of the indicated type, multiple 7–8mer 3'UTR sites, or no site. Only genes that passed our threshold for quantification (≥100 RPFs and ≥100 mRNA-Seq tags) were considered.

(C) Cumulative distributions of translational efficiency changes for the polyadenylated mRNA that remained after deleting *mir-155*. For each gene, the translational efficiency change was calculated by normalizing the RPF change by the mRNA-Seq change.

Figure S4. Sequence logos of aligned 30-nt RPFs mapping to ORFs or 5′UTRs in neutrophils.
(A) Sequence logo generated from 30-nucleotide RPFs mapping to the ORFs in neutrophils. The sequence logo is built from aligned 30-nt RPFs that mapped to the 5′-most 1% of the coding regions of detected genes (pink bar). The expected location of the P site of the ribosome is indicated (green bar). Note that in the neutrophil data set, only 2% of the RPFs mapped to the region over the start codon (compared to 14% in the activated B cells). Aligning RPFs that mapped to the 5′-most 5% of the coding regions (as in Figure 3C) would dilute the signal from ribosomes paused over the start codons, hence only RPFs that mapped to the 5′-most 1% were aligned.
(B) Sequence logo generated from 30-nucleotide RPFs mapping to the 5′UTRs in neutrophils. The sequence logo is built from aligned 30-nt RPFs that mapped to 5′UTRs of detected genes (purple bar). The expected location of the P site of the ribosome is indicated (green bar).

Figure S5. Upstream ORF translation from canonical and non-canonical start codons.
(A) Density of RPFs near the 5′ ends of uORFs that initiate with AUG, combining data from all quantified AUG-uORFs. RPF density is plotted as in Figure 1C. A schematic diagram of the ribosome is shown, with the start codon in its P site. For clarity, positions +1, +2 and +3, relative to the first base (numbered '0') of the AUG start codon, are highlighted by dashed lines (+1: blue, +2: red, +3: black). Because positions are shifted +15 nucleotides to reflect the position of the first nucleotide in the ribosome A site, a ribosome that stalls over the start codon, with the start codon in its P site, would generate an RPF whose 5′ end will map to the +3 position. The accumulation of such RPFs thus results in a strong peak at this position. The 3-nt periodicity characteristic of translating ribosomes can be clearly seen. The number of uORFs is indicated in parentheses. uORFs are annotated as described in the Methods section.

(B) Density of RPFs mapping near the 5′ ends of uORFs that initiate with non-canonical start codons. RPF density is plotted for quantified uORFs that initiate with codons that are one base-change away from the canonical AUG. Otherwise, as in panel **A**.

Note that except for AGG and AAG, all the non-canonical start codons exhibit the same characteristic peak at the +3 position, as seen with AUG-uORFs (panel **A**). For AGG and AAG, the peak appears to occur at the –1 position instead. This is because AGG and AAG are among the most commonly used second codons. Because most of the signal from RPFs mapping to AGG- and AAG-initiated uORFs appear to have derived from such cases, AGG- and AAG-initiated uORFs were excluded from the uORF annotations used in subsequent analyses.

Figure S6. Correspondence between ribosome changes, as measured by ribosome profiling, and changes in protein output, as measured by pSILAC.

(A) Correspondence between ribosome (RPF) and protein changes after deleting *mir-155*, plotting data for the 229 genes that had at least one miR-155 3'UTR site and were detected in the pSILAC experiment (proteomics-detected). The R^2 derived from Pearson's correlation is indicated.

Table S1. Alignment statistics for sequencing reads.

	B cell mRNA-Seq		B cell Ribosome Profiling	
	WT	*mir-155*$^{-/-}$	WT	*mir-155*$^{-/-}$
Raw number of reads	20,029,894	20,228,488	80,485,409	92,540,736
Uniquely mapping reads				
To genome (first-phase)	12,603,268	12,766,663	21,044,131	23,262,228
To reference transcript database (second-phase)	695,414	703,681	956,855	900,953
Total mapping uniquely	13,298,682	13,470,344	22,000,986	24,163,181
Percentage of raw reads mapping uniquely	66.4	66.6	27.3	26.1
rRNA unique matches	868,273	874,106	9,366,108	11,615,084
Percentage rRNA unique matches	6.9	6.8	44.5	50.0
Unambiguous, unique gene-mapping reads				
Total mapped reads (library size)	11,197,686	11,367,779	8,514,762	8,189,366
Percentage of all unique matches	84.2	84.4	38.7	33.9
Number of reads mapping to exons	7,694,185	7,806,937	8,235,063	7,904,395
Number of reads mapping to coding exons	5,410,118	5,490,859	7,114,588	6,774,306
Percentage coding exon reads, out of all reads mapping to exons	70.3	70.3	86.4	85.7
Number of reads mapping to initial peak	19,338	19,492	959,570	966,381
Percentage of reads mapping to initial peak, out of all coding exon reads	0.4	0.4	13.5	14.3
Number of reads mapping to 5'UTR	261,639	264,894	938,296	952,759
Number of reads mapping to annotated upstream ORFs	170,432	172,693	874,081	891,281
Percentage of 5'UTR reads mapping to annotated upstream ORFs	65.1	65.2	93.2	93.5

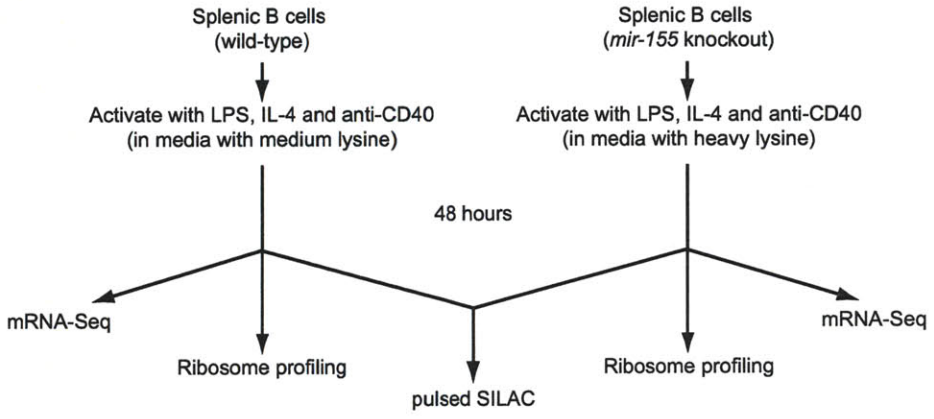
Table S2. Breakdown of codon representation in sequence logo alignment in Figure 3D.

Codon	Representation in alignment
CUG	19.0%
AUG	12.0%
GUG	7.8%
ACG	4.9%
UUG	3.9%
AUU	2.2%
GAG	2.2%
CCU	2.0%

*Only the top 8 codons are shown.

Figure S1

A


Splenic B cells
(wild-type)
Activate with LPS, IL-4 and anti-CD40
(in media with medium lysine)
Splenic B cells
(*mir-155* knockout)
Activate with LPS, IL-4 and anti-CD40
(in media with heavy lysine)
48 hours
mRNA-Seq
Ribosome profiling
pulsed SILAC
Ribosome profiling
mRNA-Seq

B

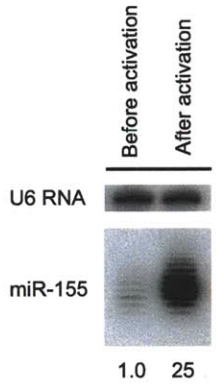

Before activation
After activation
U6 RNA
miR-155
1.0
25

Figure S2

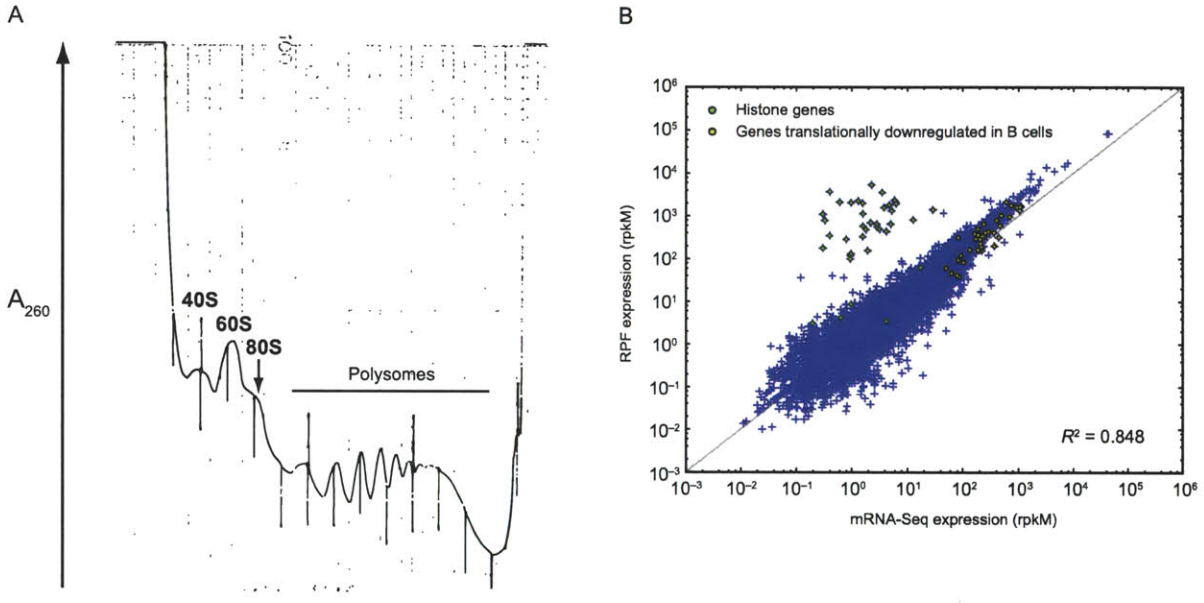

A
A_{260}
40S
60S
80S
Polysomes
B
Histone genes
Genes translationally downregulated in B cells
RPF expression (rpkM)
mRNA-Seq expression (rpkM)
R^2 = 0.848

Figure S3

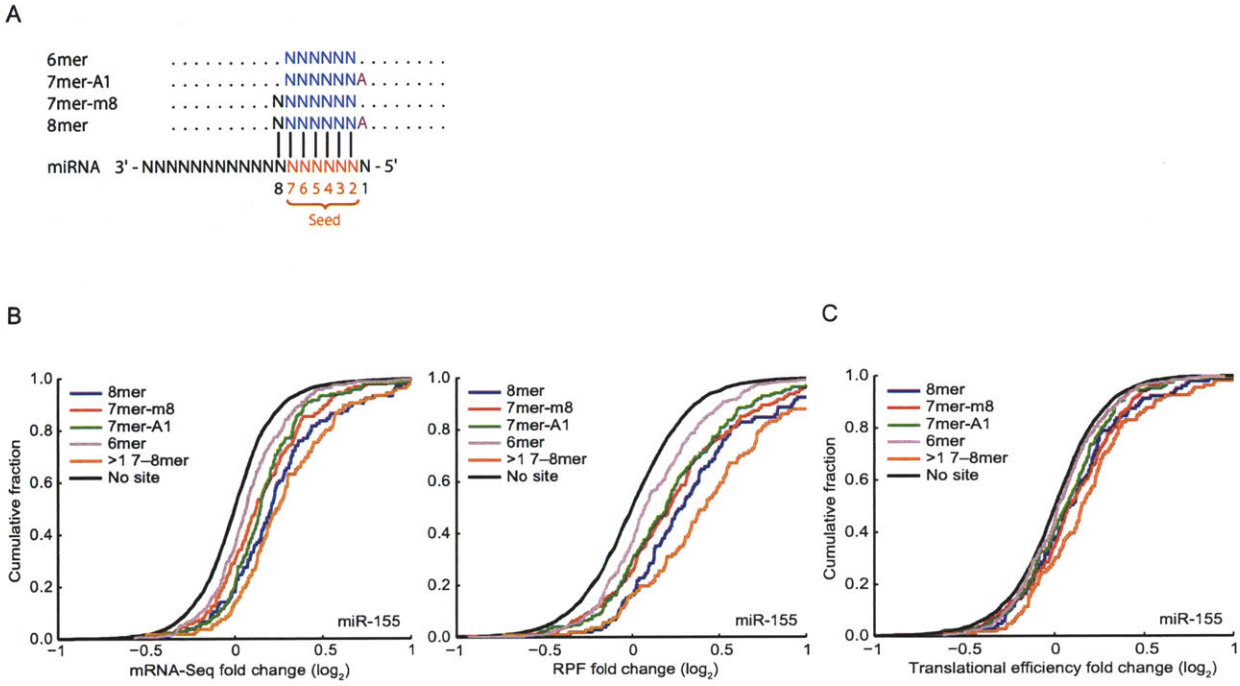

A
6mer ....NNNNNN.....
7mer-A1 ....NNNNNNA.....
7mer-m8 ....NNNNNNN.....
8mer ....NNNNNNNA.....
miRNA 3' - NNNNNNNNNNNNNNNNNNNN - 5'
8 7 6 5 4 3 2 1
Seed
B
C
8mer
7mer-m8
7mer-A1
6mer
>1 7–8mer
No site
Cumulative fraction
miR-155
mRNA-Seq fold change (log2)
RPF fold change (log2)
Translational efficiency fold change (log2)

Figure S4

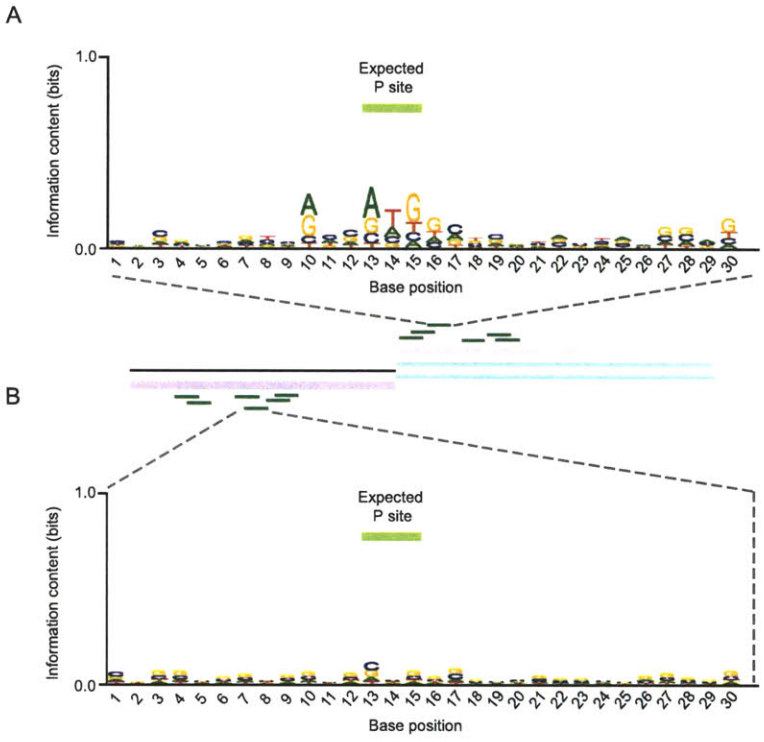

A
1.0
0.0
Information content (bits)
Expected
P site
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
Base position
B
1.0
0.0
Information content (bits)
Expected
P site
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
Base position

Figure S5

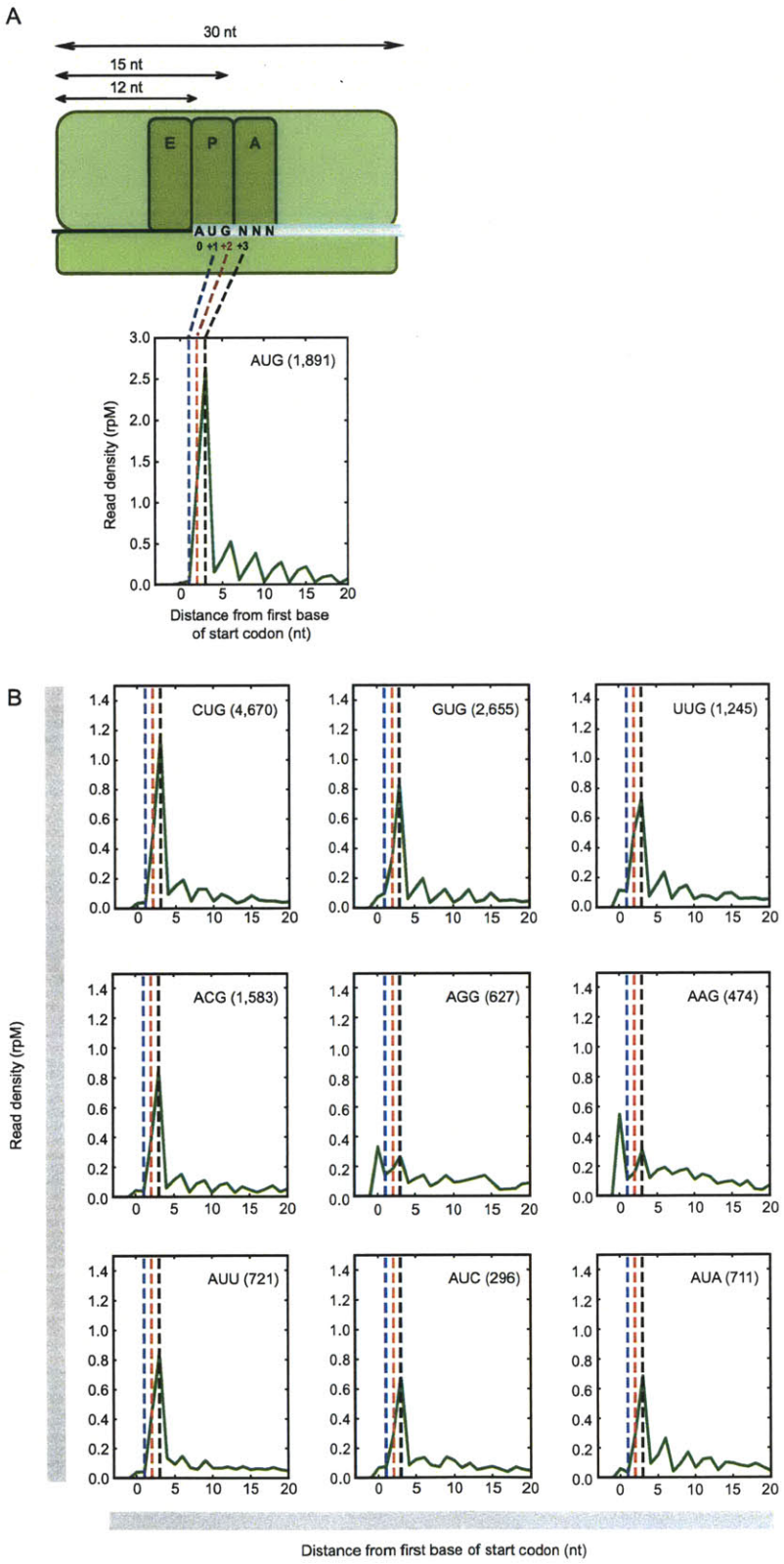
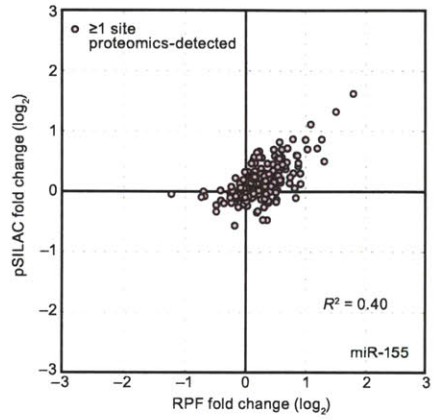

A
30 nt
15 nt
12 nt
E
P
A
AUG NNN
0 +1 +2 +3
AUG (1,891)
Read density (rpM)
Distance from first base of start codon (nt)
B
CUG (4,670)
GUG (2,655)
UUG (1,245)
ACG (1,583)
AGG (627)
AAG (474)
AUU (721)
AUC (296)
AUA (711)
Read density (rpM)
Distance from first base of start codon (nt)

Figure S6

A


≥1 site
proteomics-detected
R^2 = 0.40
miR-155
pSILAC fold change (log_2)
RPF fold change (log_2)
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

Chapter 4

Future directions

Implications for microRNA-mediated repression

In this thesis, we have shown that mammalian microRNAs (miRNAs) primarily exert their impact on target mRNAs at the mRNA level. By using ribosome profiling and mRNA-Seq in parallel, a measure of translational efficiency can be derived for thousands of genes without the need to use proteomics. This genomics-based approach also provides a level of dynamic range and robustness that cannot be matched by proteomics.

The fact that the majority of the impact is exerted at the mRNA level suggests that the inherent stability of transcripts could influence the extent of miRNA-mediated repression. For example, it is conceivable that the apparent repression of a target mRNA that is constitutively unstable would be less than one that has a longer half-life; in other words, the latter's stability maximizes the likelihood of observing the full extent of destabilization effects exerted by a given miRNA. Interestingly, a recent study that analyzed microarray data from multiple miRNA transfection data sets found a positive correlation between repression extent and target mRNA stability (Larsson et al., 2010). This suggests that target prediction algorithms could be improved if provided with accurate information of mRNA half-lives.

Although the major component of miRNA-mediated target downregulation is attributable to mRNA destabilization, there is still a component contributed by translational repression. The observed reduction in ribosome density is likely due to a block at translation initiation. Ribosome profiling would not have been able to detect mechanisms that reduce protein output without changes in ribosome density. However, through comparisons of ribosome profiling results and sample-matched proteomic data, we have also shown that it is unlikely that miRNA-mediated reduction in protein output is due to cotranslational degradation, or a combination of initiation inhibition and elongation blockage that maintains the density of ribosomes.

Repression of upstream open reading frames

The use of ribosome profiling has also enabled us to detect miRNA-mediated translational repression of upstream open reading frames (uORFs). This is an effect that proteomics approaches not tailored towards the examination of short peptides would likely miss. The finding that uORFs are repressed translationally is also consistent with miRNAs inhibiting a step during translation initiation, such as by reducing ribosome loading.

The presence of translated uORFs is thought to interfere with translation of the main coding region (Calvo et al., 2009). In addition to evidence suggesting that miRNAs are likely to reduce ribosome loading, it is tempting to consider the possibility that uORFs, being encountered first by the ribosome, could influence translational repression of the downstream coding region, potentially adding a layer of buffering capacity to miRNA repression of the main ORF. However, if so, the mechanism is unlikely to be a straightforward one. In the presence of miRNA targeting, the contexts in which the different start codons (uORF and main ORF) occur in are still encoded by the same *cis*-sequence, thus ribosomes that are loaded onto a transcript are just as likely to initiate with the same relative frequencies, at the different start codons, as before miRNA targeting occurred. This would be the case even if the number of ribosomes loaded decreases. It follows then that for uORF translation to modulate translational repression of the main ORF, the dwell time of the ribosome, or its subunits, on uORFs must change according to the presence/absence of miRNA targeting. This change could be mediated by *trans* factors, such as a different composition of scanning small subunits. In mouse embryonic stem cells, initiating ribosomes have a propensity to recognize non-canonical start codons in 5′UTRs, a phenomenon that was not observed in the main ORF, even when alternative start codons are present within the main ORF (Nicholas Ingolia, personal communication). This could be mediated by a switch in affinity for start codon recognition, or the presence of two different populations of scanning small subunits. Regardless of the mechanism, such a difference could provide the basis for differential dwell times on uORFs versus the main ORF, and thus impart the potential for translated uORFs to modulate miRNA-mediated translational

repression of the main coding region. This can be directly tested by examining miRNA repression levels under conditions that facilitate different extents of uORF translation. A plausible experimental condition is that of nutrient starvation, which has been shown to result in increased uORF translation in yeast (Ingolia et al., 2009), though this has not yet been shown in mammalian systems.

Repression mediated by sites in the open reading frame

Sites in the 3′UTR are typically more effective than sites in the ORF (Bartel, 2009). It is believed that the passage of translating ribosomes reduces the likelihood, or duration, of silencing complex occupancy that is required to mediate repression (Grimson et al., 2007). This has led to the idea that ORF sites should be more effective in messages that are inefficiently translated.

Theoretically, ribosome profiling — in parallel with mRNA-Seq — enables the quantification of translational efficiency for each gene and should allow us to address this question directly. However, we did not observe a relationship between translational efficiency and ORF-site efficacy, whether in the neutrophils or B cells. It is possible that the weak repression mediated by ORF targeting, coupled with experimental noise, precludes the observation of a statistically significant trend. It could also be that additional confounding factors mask the presence of a weak trend. One possible such factor might be the physical proximity of a target site to the poly(A)-binding protein (PABP), which has been reported to assist in miRNA-mediated repression (Fabian et al., 2009; Zekri et al., 2009). This would suggest that a subset of ORF targets — those with 3′UTR-proximal ORF sites — are likely to be much more responsive than most potential ORF targets; if so, a relationship between translational efficiency and ORF-site efficacy might only be detected if, during data analysis, these responsive genes were isolated from all other genes with ORF sites. Interestingly, sites further along in the open reading frame have been found to be more effective, but this was only observed after combining microarray data from multiple miRNA transfections (Garcia et al., *In press*), underscoring the difficulty in delineating rules of ORF targeting from single data sets.

Another analysis that could benefit from having increased amounts of data is the probing of uORF translational repression by sites in the ORF, which could serve as effective 3′UTR sites if ribosome passage through the main coding region is very much reduced by uORF translation. Concurrently, this may also amplify mRNA destabilization effects for targets with both 3′UTR and ORF sites, compared to targets with 3′UTR sites alone. However, if sites in the ORF need to be far away from the start codon in order to be effective, these analyses would similarly be confounded by the same reasoning as previously described, unless multiple data sets can be combined to yield enough responsive target genes to observe statistically significant trends.

Primary event of miRNA targeting: first translational repression, or first decay?

One question that has arisen from the genome-wide studies that found mRNA destabilization to be the major component of miRNA-mediated repression is whether the destabilization effect is triggered by an initial translational repression.

In yeast, the disruption of certain translation initiation factors has been shown to destabilize mRNAs (Schwartz and Parker, 1999), leading to the idea that efficient translation prevents mRNAs from getting diverted to decay pathways. However, this does not necessarily imply causation because disrupting initiation factors in the cap binding complex, as this study did, would likely disrupt circularization of the mRNA, which could in turn have a negative impact on both translation efficiency and mRNA stability. In addition, examples abound in eukaryotes in which translationally repressed mRNAs are not automatically destabilized. In neurons, mRNAs in dendrites are translationally induced upon neuronal activation, and these mRNAs are believed to be translationally silent, yet stable, during transport from the cell body to the dendrites (Giorgi et al., 2007). In cultured cells, ~30% of mRNAs are present as free messenger ribonucleoprotein (mRNP) particles[1] (Mathews et al., 2007). Interestingly, although it was originally thought that only messages that are not associated with ribosomes can be

[1] Which should not exist at all, if mRNAs that are not associated with polysomes are immediately degraded.

channeled into the decay pathways (Coller and Parker, 2004; Parker and Sheth, 2007), it has recently been shown in yeast that decapped mRNAs are associated with polysomes and 5′→3′ decay takes place even as the last ribosomes are completing translation of the mRNA (Hu et al., 2009). Thus, the presence of ribosomes does not necessarily protect a message from degradation.

The converse scenario, in which miRNA targeting triggers deadenylation, which then leads to translational repression before any decay of the mRNA body, could also be best explained by the need to maintain the 'closed loop' conformation. PABP is an integral component of the 'closed loop' model. While only one PABP can interact with eIF4G at any point in time, having a longer tail (and thus more attached PABPs) raises the effective concentration of PABP, and increases the likelihood that the 'closed loop' is maintained, which in turn increases the efficiency of translation initiation. This benefit would be compromised by deadenylation.

Because breaking the 'closed loop' conformation is likely to affect both translational efficiency and mRNA stability (Gallie, 1991), pinpointing the exact component that is most immediately targeted by miRNAs would require more inventive approaches. Moreover, because any targeted factor is likely to disrupt the 'closed loop', whether translational repression or deadenylation comes first may eventually be a matter of academic definition, in terms of molecular outcome — for example, the suggested disruption of the eIF4G:PABP interaction by GW182 (Fabian et al., 2009; Zekri et al., 2009) would be a contributing factor towards decircularization, rather than translational repression *per se*, not least because the reported GW182:PABP-assisted recruitment of deadenylases upon miRNA targeting would also accelerate decay of the target mRNA (Fabian et al., 2009; Zekri et al., 2009).

Stress and the balance between translational repression and mRNA destabilization

In Chapter 2, we saw that in the initial data sets from miRNA transfections into HeLa cells, and the *in vitro* differentiated neutrophils, translational repression contributed 11–16% of the overall microRNA-

mediated repression. Although the major impact was still exerted at the mRNA level in activated B cells (Chapter 3), translational repression made up a larger component (~29%). What could explain this difference?

Intriguingly, activated B cells are believed to be undergoing a physiological form of the unfolded protein response (UPR)(Brewer and Hendershot, 2005). Massive upregulation of immunoglobulin synthesis is thought to bring about sustained stress in the endoplasmic reticulum (ER), yet B cells are able to avoid apoptosis, which is the typical consequence of such prolonged stress (Tabas and Ron, 2011). As mentioned in Chapter 1, an integral component of the UPR is PERK, whose activation leads to global shut-down of translation by reducing the availability of the eIF2:GTP:Met-tRNA$_i$ ternary complex (eIF2-TC). In addition to PERK, two other branches make up the UPR in mammalian cells (Todd et al., 2008), one of which is the IRE1 pathway, which is also conserved in yeast. IRE1 is a transmembrane protein on the ER surface that when activated, dimerizes to cleave the XBP1 mRNA at two locations. The XBP1 mRNA normally makes XBP1u protein; upon cleavage and ligation of its 5′- and 3′-most fragments, the newly spliced mRNA now makes XBP1s protein, a transcription factor that upregulates other UPR-related genes via UPR elements (UPREs), which are *cis*-acting promoter sequences. ATF4, the transcription factor that is paradoxically synthesized due to PERK activation (see Chapter 1) also activates UPRE-promoters. The last branch of the UPR is triggered by ATF6, another ER transmembrane protein. When activated, ATF6 trafficks to the Golgi where it is cleaved to produce yet another transcription factor. This transcription factor upregulates genes with ER stress response elements (ERSEs). Together with the global translational downregulation, the transcriptional program driven by UPREs and ERSEs coordinate an integrated stress response.

XBP1 is essential for the differentiation of activated B cells into plasma cells (Reimold et al., 2001), and the production of XBP1s is upregulated only when the IgM locus is present, leading to the idea that intensive immunoglobulin synthesis triggers ER stress in activated B cells (Iwakoshi et al., 2003). PERK$^{-/-}$ mice, however, do not show any deficiency in B cell function (Gass et al., 2008) and CHOP, a gene that is normally induced in a PERK-dependent fashion is not expressed when B lymphoma

cells were induced to differentiate (Gass et al., 2002). This led to the hypothesis that a form of physiological UPR is effected in activated B cells. Activated B cells eventually need to synthesize large amounts of antibodies, and even before that, would need to upregulate their secretory capacity; thus it may not be productive to activate the PERK branch of the UPR.

Knowing that the activated B cells are likely undergoing some form of stress, could this stressful condition result in a slight shift in balance between the mRNA destabilization component and the translational repression component (Leung and Sharp, 2010)? The biological importance of miRNAs during stressful conditions is evident from certain miRNA-knockout organisms, in which observable phenotypes are not apparent until the animal is stressed (Li et al., 2009; van Rooij et al., 2007; Xu et al., 2003). Might a shift in balance between the two components be able to contribute to differential biological outcomes during stress? Although the model of miRNAs disrupting the 'closed loop' conformation of target mRNAs awaits confirmation, it is the model that is most consistent with currently available evidence (Huntzinger and Izaurralde, 2011), and thus will be used as a framework for discussion in the following sections.

Stress and its potential effect on mRNA stability

It was recently reported that *lin-4* may repress *lin-14* via different modes when worms are grown in different conditions. In worms grown under normal conditions, *lin-4* represses *lin-14* mostly at the mRNA level (Bagga et al., 2005); under nutrient-deprived conditions, LIN-14 protein levels are still reduced by the same amount but *lin-14* mRNA is stabilized, suggesting that during nutrient deprivation, the reduction of LIN-14 protein output is primarily mediated at the translation level, though it has yet to be convincingly shown that this translational repression is due to the action of *lin-4* alone (Holtz and

Pasquinelli, 2009). In addition, two other miRNA:target pairs tested did not exhibit the sensitivity of *lin-4:lin-14* towards starvation conditions[2].

Nevertheless, it is possible that mRNA stability could be altered by stress conditions, and in turn influence the outcome of miRNA-mediated repression. Several reports have shown that the stability of short-lived mRNAs increases upon various stresses (Gowrishankar et al., 2005; Gowrishankar et al., 2006). In particular, it was shown in yeast that the activity of certain deadenylases was reduced upon osmotic stress, heat shock and glucose deprivation (Hilgers et al., 2006). If this observation holds true in mammals, then it is conceivable that even when a target mRNA is decircularized upon miRNA targeting, it may not be destabilized to the same extent as under normal conditions when the activity of decay enzymes is not impaired.

Even if the activity of the decay machinery were not compromised, reduced ribosome loading during stressful conditions might still impact mRNA stability. The release factor eRF3 interacts directly with PABP, and this interaction minimizes the multimerization of PABP monomers on the poly(A) tail *in vitro* (Hoshino et al., 1999). Because PABP multimerization contributes to the protection of the poly(A) tail, reducing multimerization could expedite access of the decay machinery to the poly(A) tail, and in turn, link translation termination to mRNA decay. Linking translation and mRNA turnover would be a potential mechanism for a 'clock' that limits the lifespan of polyadenylated mRNA. Interestingly, this mechanism is reminiscent of Sussman's 'ticketing' hypothesis, in which the 5′ region of the mRNA serves a role that could be played by what we know today as the poly(A) tail (Sussman, 1970). Extrapolating from this observation, it is possible that when global translation is reduced under stressful conditions, there would be fewer terminating ribosomes, with concomitantly increased stability of polyadenylated transcripts. Interestingly, in a genome-wide study that simultaneously examined mRNA

[2] Because nutrient deprivation had such a large influence on the mode of reduction of LIN-14 protein output, it was speculated that differential growth conditions might have contributed to the reported discrepancies between the initial *lin-4:lin-14* studies that concluded miRNAs repress their targets translationally (Wightman et al, Olsen and Ambros, 1997) and the later study that showed *lin-4* primarily reducing *lin-14* mRNA levels (Bagga et al, 2005).

stability and translational status when mammalian cells were subjected to ER stress, mRNAs that were stabilized in response to stress were preferentially subject to translational repression (Kawai et al., 2004).

It is important to note that a general increase in mRNA stability would concurrently lead to more detectable translational repression by the methods used in this thesis. This is because the longer the mRNA is around after miRNA targeting, the higher the likelihood that the reduced ribosome loading that ensues, upon disruption of the 'closed loop', would be reflected as a reduction in translational efficiency. Conversely, if an mRNA is immediately degraded upon miRNA targeting, no effect on ribosome loading would be seen because the state at which there were fewer ribosomes on the target mRNA would not have existed. Experiments that allow parallel measurements of translational efficiency and mRNA stability in the absence/presence of stress would be able to address the above-described hypothesis.

Stress and its potential effect on translational efficiency

On the other hand, if an mRNA that has been decircularized is still well-translated, then an overall increase in mRNA stability would not lead to an apparent increase in the amount of translational repression, as measured by ribosome profiling and mRNA-Seq. This, however, leads to another potential explanation for the discrepancy in contributions from the translational repression component between activated B cells, and the actively-dividing HeLa cells and neutrophils.

The exact mechanism by which 'closed loop' formation stimulates translation is unclear, but the best-supported mechanism is the recruitment of ribosomal subunits to the mRNA (Sachs et al., 1997). This effect is thought to play more of a role when translation factors and/or ribosomes are limiting (Proweller and Butler, 1997). Ribosome production is controlled not just by the coordinate regulation of ribosomal protein expression via the 5′-terminal oligopyrimidine tract (5′TOP), but also in terms of ribosomal subunit production. Knocking down mRNAs encoding individual ribosomal proteins leads to a concomitant decrease in the abundance of all proteins from the same subunit (Robledo et al., 2008). This tight coordination is possible because ribosomal proteins are rapidly degraded if subunit assembly is

somehow impaired (Warner, 1977). Cells that are actively dividing, such as HeLa cells and neutrophils, are constantly synthesizing new ribosomes; in activated B cells, in which the translation of ribosomal proteins is downregulated (see Figure 1E in Chapter 3), the concentration of ribosomal subunits may well have dropped enough such that productive translation initiation becomes more dependent on circularization. In other words, if 'closed loop' formation stimulates translation initiation by promoting subunit recruitment, then it is conceivable that disrupting the 'closed loop' is likely to be more detrimental under conditions of low ribosomal subunit concentrations, when the lack of stimulatory activity is less likely to be compensated for by mass action. Conversely, in actively-dividing cells that are primed for growth, ribosome loading might still occur for a decircularized mRNA, albeit at a lower rate, by virtue of a high concentration of available ribosomal subunits.

Thus, provided that the pre-existing ribosomal subunits are not present in concentrations high enough to negate this possibility, this could explain the observed discrepancy between the B cells and the actively-dividing HeLa cells and neutrophils. This hypothesis predicts that lowering the concentration of ribosomal subunits such that they become limiting in actively-dividing cells would increase the contribution of translational repression towards the overall miRNA-mediated repression. Because the production of ribosomal subunits is so tightly coordinated, this prediction can be explicitly tested by knocking down a few ribosomal proteins in HeLa cells to lower ribosomal subunit concentrations artificially.

It is important to note that even if ribosomal subunits were not limiting, a higher translational repression component could also be observed if translation factors, such as initiation factors, were limiting. Interestingly, among the translationally downregulated genes in the activated B cells is *Eif3f* (see Figure 1E in Chapter 3). eIF3F, incidentally, is one of the three subunits, of human eIF3, that are validated 5′TOP mRNAs (Iadevaia et al., 2008). Because 5′TOP mRNAs are not well-translated in activated B cells, it is possible that eIF3 concentration in these cells is lower than that in actively dividing cells. One of the ways by which eIF3 is thought to facilitate translation initiation is by binding the 40S subunit and preventing its association with the 60S subunit (Hinnebusch, 2006). Lowered eIF3

concentrations could thus lead to fewer free 40S and 60S subunits available for initiating new rounds of translation, and might also explain the preponderance of 80S ribosomes[3] in the activated B cells (Figure 1A and 1B in Chapter 3). Thus, even if the pre-existing pool of ribosomes is not degraded after the cessation of new ribosomal subunit production, free subunits could still be sequestered due to lowered eIF3 concentrations. As such, knocking down eIF3, in conjunction with ribosomal protein knockdowns, might be more effective in testing the hypothesis described in the previous paragraphs.

More translational repression, or more decay — does it matter?

Regardless of the exact mechanism of repression, it is now clear that miRNAs repress their targets by a combination of translational repression and mRNA destabilization. There has been much debate over which component plays a more important role, but if either (or a combination of both) eventually lead to the same outcome, i.e. repression of the target, does it really matter which component is the major one? To answer this question, it would perhaps help to think in terms of extreme examples, such as a target that is only translationally repressed (target A) and, another that is only destabilized (target B).

One obvious difference between targets A and B is that pure translational repression does not deplete the existing pool of target mRNAs (Leung and Sharp, 2010). Thus, for target A, it would be possible to mount a swift recovery of protein synthesis should the need arise, whereas for target B, any recovery would be slower as new rounds of transcription would be required to replenish the reduced pool of target mRNAs. To achieve a swift recovery, the function of the silencing complex has to be concurrently compromised. This scenario would be similar to the reported regulation of CAT-1 mRNA by miR-122 in liver cells, in which translationally repressed CAT-1 mRNA is moved out of P bodies, and back into the translating pool, because the binding of the HuR protein to the CAT-1 3′UTR compromises the repressive action of miR-122 (Bhattacharyya et al., 2006).

[3] In this case, these 80S ribosomes would be 'empty' monosomes.

In the above example, the HuR protein is exported from the nucleus to the cytoplasm during stress, and a "recovery" of CAT-1 protein synthesis occurs during stress[4]. However, because stress in general induces global translation blockage, more biological impact could be exerted if recovery were to take place right after a stressful period, when it could capitalize on the ensuing restoration of translation. Therefore, it might be interesting to look for targets undergoing this type of regulation before/after stress. Moreover, in anticipation that the function of the silencing complex should be simultaneously downregulated, the search could be narrowed by looking for miRNAs whose function/expression may be compromised after various forms of stress. One way to approach this would be to look for RNA binding proteins whose 3′UTR binding ability is regulated by stress[5], and then examine the 3′UTRs of bound mRNAs for target sites of concurrently expressed miRNAs. Another good starting point might be to look for miRNAs with potential ERSE/UPRE-promoters, on the basis that because these miRNAs are upregulated to cope with stress, they would likely be downregulated when the stress is over.

Another scenario, perhaps less obvious but no less important, can also be illustrated using the framework provided by targets A and B. In most cases of conserved miRNA targeting — whether they are switch interactions or tuning interactions — the maintenance of target gene repression is paramount. In such a scenario, downregulation of target A would take place purely by translational repression; this relies on target A's inherent turnover rate to 'clear' the target transcripts. In contrast, target B would be actively destabilized by the miRNA. Should the prevailing conditions change such that the sole mechanism employed (translational repression or mRNA destabilization) is impaired, repression would be compromised, potentially leading to detrimental outcomes.

In contrast, a target that can be downregulated by either mechanism would allow repression to be maintained. For example, the GW182:PABP interaction was reported to facilitate the recruitment of the CCR4-NOT deadenylase complex to the targeted transcript (Fabian et al., 2009; Zekri et al., 2009). As

[4] In fact, CAT-1 mRNA has an internal ribosome entry site (IRES) that can bypass cap-dependent translation (see Chapter 1) but translation initiated from this IRES does not begin until eIF2α phosphorylation has taken place, which occurs after this initial HuR-mediated recovery.

[5] Taking the lead from HuR and miR-122.

described in the earlier section, conditions of stress might affect the activity of deadenylases (Gowrishankar et al., 2005; Gowrishankar et al., 2006; Hilgers et al., 2006). If so, the extent to which the mRNA destabilization component contributes to the overall repression could vary depending on the activity of the decay machinery. Should deadenylase activity be compromised during stress conditions, the shift in balance towards translational repression, by virtue of breaking the 'closed loop' conformation and reducing ribosome loading, serves as a back-up mechanism to maintain repression. In effect, targeting both stability and translation (for example, by disrupting the 'closed loop') would buffer repression levels against variable conditions; this would in turn reinforce the 'regulatory network buffering' role that miRNAs are believed to play in the cell (Bartel, 2009), and would be more advantageous than a repression system that takes place only via one mechanism. Another interesting question is whether the extent of overall repression might be different if/when the balance between the two components shifts; this would be an important point to address because a difference in magnitude is likely to have greater biological consequences than a difference in balance.

Concluding remarks

It is now better appreciated that post-transcriptional regulation offers a more flexible and rapid way of controlling gene expression, compared to transcriptional control. The buffering capacity afforded by miRNAs, when operating with increased transcription, also serves to smooth out gene expression (Bartel and Chen, 2004), considering that transcription tends to occur in bursts (Suter et al., 2011), which can lead to more choppy expression levels. The fact that so many miRNA:target interactions have been selectively maintained over evolution, despite the subtle changes they impart, suggests that maintaining optimal levels of protein expression is of great importance to animal fitness. The premium placed on such tight tolerances is one of the more fascinating findings of the miRNA field.

When we look back at Jacob and Monod's gene regulation models and Sussman's 'ticketing' hypothesis, the core features have proven to be prescient even though the modern elaborations are more complex, and may even be quite different, than originally envisioned. Thus, biology continues to surprise

us, and it is wise to expect the unexpected. At the same time, however, it is prudent to remember that we are standing on the giant shoulders of past discoveries and insights. As Louis Pasteur once said, "Fortune favors the prepared mind" — being aware of the huge strides that have already been made before us can be a valuable part of our arsenal in anticipation of the unexpected.

Acknowledgements

I thank David Bartel and Muhammed Yildirim for the many stimulating discussions that have contributed to the development of ideas presented in this chapter. I also thank Nicholas Ingolia for sharing unpublished data.

References

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A.E. (2005). Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. Cell *122*, 553-563.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Bartel, D.P., and Chen, C.Z. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet *5*, 396-400.

Bhattacharyya, S.N., Habermacher, R., Martine, U., Closs, E.I., and Filipowicz, W. (2006). Relief of microRNA-mediated translational repression in human cells subjected to stress. Cell *125*, 1111-1124.

Brewer, J.W., and Hendershot, L.M. (2005). Building an antibody factory: a job for the unfolded protein response. Nat Immunol *6*, 23-29.

Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci U S A *106*, 7507-7512.

Coller, J., and Parker, R. (2004). Eukaryotic mRNA decapping. Annu Rev Biochem *73*, 861-890.

Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A., *et al.* (2009). Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. Mol Cell *35*, 868-880.

Gallie, D.R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Dev *5*, 2108-2116.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. (*In press*). Weak Seed-Pairing Stability and High Target-Site Abundance Decrease the Proficiency of *lsy-6* and Other miRNAs Nat Struct Mol Biol.

Gass, J.N., Gifford, N.M., and Brewer, J.W. (2002). Activation of an unfolded protein response during differentiation of antibody-secreting B cells. J Biol Chem *277*, 49047-49054.

Gass, J.N., Jiang, H.Y., Wek, R.C., and Brewer, J.W. (2008). The unfolded protein response of B-lymphocytes: PERK-independent development of antibody-secreting cells. Mol Immunol *45*, 1035-1043.

Giorgi, C., Yeo, G.W., Stone, M.E., Katz, D.B., Burge, C., Turrigiano, G., and Moore, M.J. (2007). The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. Cell *130*, 179-191.

Gowrishankar, G., Winzen, R., Bollig, F., Ghebremedhin, B., Redich, N., Ritter, B., Resch, K., Kracht, M., and Holtmann, H. (2005). Inhibition of mRNA deadenylation and degradation by ultraviolet light. Biol Chem *386*, 1287-1293.

Gowrishankar, G., Winzen, R., Dittrich-Breiholz, O., Redich, N., Kracht, M., and Holtmann, H. (2006). Inhibition of mRNA deadenylation and degradation by different types of cell stress. Biol Chem *387*, 323-327.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell *27*, 91-105.

Hilgers, V., Teixeira, D., and Parker, R. (2006). Translation-independent inhibition of mRNA deadenylation during stress in Saccharomyces cerevisiae. RNA *12*, 1835-1845.

Hinnebusch, A.G. (2006). eIF3: a versatile scaffold for translation initiation complexes. Trends Biochem Sci *31*, 553-562.

Holtz, J., and Pasquinelli, A.E. (2009). Uncoupling of lin-14 mRNA and protein repression by nutrient deprivation in Caenorhabditis elegans. RNA *15*, 400-405.

Hoshino, S., Imai, M., Kobayashi, T., Uchida, N., and Katada, T. (1999). The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3'-Poly(A) tail of mRNA. Direct association of erf3/GSPT with polyadenylate-binding protein. J Biol Chem *274*, 16677-16680.

Hu, W., Sweet, T.J., Chamnongpol, S., Baker, K.E., and Coller, J. (2009). Co-translational mRNA decay in Saccharomyces cerevisiae. Nature *461*, 225-229.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet *12*, 99-110.

Iadevaia, V., Caldarola, S., Tino, E., Amaldi, F., and Loreni, F. (2008). All translation elongation factors and the e, f, and h subunits of translation initiation factor 3 are encoded by 5'-terminal oligopyrimidine (TOP) mRNAs. RNA *14*, 1730-1736.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Iwakoshi, N.N., Lee, A.H., Vallabhajosyula, P., Otipoby, K.L., Rajewsky, K., and Glimcher, L.H. (2003). Plasma cell differentiation and the unfolded protein response intersect at the transcription factor XBP-1. Nat Immunol *4*, 321-329.

Kawai, T., Fan, J., Mazan-Mamczarz, K., and Gorospe, M. (2004). Global mRNA stabilization preferentially linked to translational repression during the endoplasmic reticulum stress response. Mol Cell Biol *24*, 6773-6787.

Larsson, E., Sander, C., and Marks, D. (2010). mRNA turnover rate limits siRNA and microRNA efficacy. Mol Syst Biol *6*, 433.

Leung, A.K., and Sharp, P.A. (2010). MicroRNA functions in stress responses. Mol Cell *40*, 205-215.

Li, X., Cassidy, J.J., Reinke, C.A., Fischboeck, S., and Carthew, R.W. (2009). A microRNA imparts robustness against environmental fluctuation during development. Cell *137*, 273-282.

Mathews, M., Sonenberg, N., and Hershey, J.W.B. (2007). Translational control in biology and medicine, 3rd edn (Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press).

Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. Mol Cell *25*, 635-646.

Proweller, A., and Butler, J.S. (1997). Ribosome concentration contributes to discrimination against poly(A)- mRNA during translation initiation in Saccharomyces cerevisiae. J Biol Chem *272*, 6004-6010.

Reimold, A.M., Iwakoshi, N.N., Manis, J., Vallabhajosyula, P., Szomolanyi-Tsuda, E., Gravallese, E.M., Friend, D., Grusby, M.J., Alt, F., and Glimcher, L.H. (2001). Plasma cell differentiation requires the transcription factor XBP-1. Nature *412*, 300-307.

Robledo, S., Idol, R.A., Crimmins, D.L., Ladenson, J.H., Mason, P.J., and Bessler, M. (2008). The role of human ribosomal proteins in the maturation of rRNA and ribosome production. RNA *14*, 1918-1929.

Sachs, A.B., Sarnow, P., and Hentze, M.W. (1997). Starting at the beginning, middle, and end: translation initiation in eukaryotes. Cell *89*, 831-838.

Schwartz, D.C., and Parker, R. (1999). Mutations in translation initiation factors lead to increased rates of deadenylation and decapping of mRNAs in Saccharomyces cerevisiae. Mol Cell Biol *19*, 5247-5256.

Sussman, M. (1970). Model for quantitative and qualitative control of mRNA translation in eukaryotes. Nature *225*, 1245-1246.

Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. Science *332*, 472-474.

Tabas, I., and Ron, D. (2011). Integrating the mechanisms of apoptosis induced by endoplasmic reticulum stress. Nat Cell Biol *13*, 184-190.

Todd, D.J., Lee, A.H., and Glimcher, L.H. (2008). The endoplasmic reticulum stress response in immunity and autoimmunity. Nat Rev Immunol *8*, 663-674.

van Rooij, E., Sutherland, L.B., Qi, X., Richardson, J.A., Hill, J., and Olson, E.N. (2007). Control of stress-dependent cardiac growth and gene expression by a microRNA. Science *316*, 575-579.

Warner, J.R. (1977). In the absence of ribosomal RNA synthesis, the ribosomal proteins of HeLa cells are synthesized normally and degraded rapidly. J Mol Biol *115*, 315-333.

Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A. (2003). The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. Curr Biol *13*, 790-795.

Zekri, L., Huntzinger, E., Heimstadt, S., and Izaurralde, E. (2009). The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. Mol Cell Biol *29*, 6220-6231.

Appendix

The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs

Katsutomo Okamura[1], Wei-Jen Chung[1], J. Graham Ruby[2], Huili Guo[2], David P. Bartel[2] and Eric C. Lai[1]

[1]Sloan-Kettering Institute, Department of Developmental Biology, 521 Rockefeller Research Laboratories, 1275 York Avenue, Box 252, New York, New York 10065, USA.
[2]Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, and Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA.

H.G. performed initial hpRNA northern analysis. J.G.R. identified hp-CG4068 and hpRNA1. W.-J.C. performed the EINVERTED analysis and identified the additional hpRNA loci and their targets. K.O designed and carried out all other experiments. All authors contributed to the preparation of the manuscript.

Published as:
Okamura, K., Chung, W.-J., Ruby, G. R., Guo, H., Bartel, D. P., Lai, E. C. (2008) The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature 453: 803-806

Vol 453|5 June 2008|doi:10.1038/nature07015

LETTERS

The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs

Katsutomo Okamura[1], Wei-Jen Chung[1], J. Graham Ruby[2], Huili Guo[2], David P. Bartel[2] & Eric C. Lai[1]

In contrast to microRNAs and Piwi-associated RNAs, short interfering RNAs (siRNAs) are seemingly dispensable for host-directed gene regulation in *Drosophila*. This notion is based on the fact that mutants lacking the core siRNA-generating enzyme Dicer-2 or the predominant siRNA effector Argonaute 2 are viable, fertile and of relatively normal morphology[1,2]. Moreover, endogenous *Drosophila* siRNAs have not yet been identified. Here we report that siRNAs derived from long hairpin RNA genes (hpRNAs) programme Slicer complexes that can repress endogenous target transcripts. The *Drosophila* hpRNA pathway is a hybrid mechanism that combines canonical RNA interference factors (Dicer-2, Hen1 (known as CG12367) and Argonaute 2) with a canonical microRNA factor (Loquacious) to generate ~21-nucleotide siRNAs. These novel regulatory RNAs reveal unexpected complexity in the sorting of small RNAs, and open a window onto the biological usage of endogenous RNA interference in *Drosophila*.

Artificial, long-inverted repeat transcripts are efficiently processed by a Dicer-2 (Dcr-2)/Argonaute 2 (AGO2)-driven RNA interference (RNAi) pathway in transgenic *Drosophila*[1,3]. We hypothesized that this might reflect the existence of an endogenous pathway that accepts long, inverted repeat transcripts. To test this idea, we searched for inverted repeats using EINVERTED[4] and selected putative hairpins containing mapped small RNA reads (see Methods). Out of 8,132 candidate regions, most consisted of the terminal inverted repeats of individual transposable elements or long terminal repeats of tandem inverted transposable elements. The remaining loci corresponded to inverted tandem duplications of messenger RNA- or transfer RNA-encoding genes, a microRNA (miRNA) gene (*mir-997*), a novel tandem pair of short hairpins (*chou39-1* and *chou39-2*, Supplementary Fig. 1), and a handful of single-gene annotations and unannotated regions.

We analysed the size distribution of cloned RNAs from all of the non-transposable-element EINVERTED hits. Although these mostly exhibited a broad length distribution across the ~18–26-nucleotide cloning range, indicative of degradation fragments (Supplementary Fig. 1b), seven genomic regions specifically generated 21–22-nucleotide RNAs (Supplementary Fig. 1c–f). These included genes annotated as CG18854, CG32207, CR32205 and pncr009 (also known as pncr009:3L), a series of 20 repeats that partially overlap the 3′ untranslated region (UTR) of CG4068, and an intergenic region adjacent to CG4770 (Supplementary Figs 2–7). Except for CG4068, the coding potential of all of these loci is limited[5,6]. Still, we chose to introduce an 'hp' prefix to these six loci to distinguish the small-RNA-generating hairpins ('hpRNAs') from the potential protein-encoding segments of these transcripts.

The hpRNA hairpins were collectively much longer than typical animal pre-miRNAs, and several were even longer than plant miRNAs[7]. All hpRNA loci produced dominant small RNAs that presented duplexes with 2-nucleotide 3′ overhangs, implying RNase III processing (Fig. 1 and Supplementary Figs 2–7). Such an origin was more evident with the hp-CG18854 and hp-CG4068 hairpins, from which many consecutive, phased, small RNA duplexes were cloned (Fig. 1). The 20 tandem repeats at the hp-CG4068 locus were suggestive of local duplications, and created potential for a vast array of higher-order hairpin conformations (Fig. 1b and Supplementary Fig. 3). In addition, hp-CR32205, hp-pncr009 and hp-CG32207 were related in sequence and located within a 70-kb interval (Supplementary Figs 5–8). Thus, hpRNAs, like miRNAs, can apparently evolve as local genomic duplications.

We probed the consequences of dsRNA-mediated knockdown of candidate factors on hpRNA biogenesis. We first confirmed the potency of these knockdowns by analysing Bantam, the pre-miRNA and/or mature miRNA of which were sensitive to Drosha, Pasha, Dcr-1, Loquacious (Loqs), Exportin-5 (Exp5, also known as Ranbp21) and Argonaute 1 (AGO1), as expected (Fig. 2a). The behaviour of hp-CG4068B/D/G and hp-CG18854A contrasted sharply with that of bantam (Fig. 2a and Supplementary Fig. 10a). Consistent with their apparent derivation from phased cleavage of long inverted repeats, their processing was unaffected by Dcr-1 depletion, but was strongly dependent on Dcr-2. In addition, ~21-nucleotide hpRNA products were markedly reduced when AGO2 was depleted. Incidentally, the Dcr-2/AGO2-dependent accumulation of ~21-nucleotide (siRNA) and ~42-nucleotide (terminal loop) hp-CG4068D isoforms (Fig. 2a) provided evidence for the *in vivo* processing of both 'single-repeat' and 'double-repeat' (or higher-order) forms of the hp-CG4068 hairpins (Fig. 1b).

Several other aspects of hpRNA biogenesis deserve mention. First, we were surprised that hpRNA processing was very strongly dependent on the Dcr-1-cofactor Loqs. This was especially unexpected in light of the recent realization that the *loqs* null condition only mildly compromises the maturation of many miRNAs[8], such as Bantam (Fig. 2a). Second, mature hpRNA products declined reproducibly in AGO1-deficient cells, which suggested the possible involvement of both AGO proteins in hpRNA biogenesis and/or function. Third, knockdown of Dcr-2, AGO2 and, to a lesser extent, AGO1 resulted in a ladder of hybridizing bands consistent with impaired hairpin processing (Fig. 2a and Supplementary Fig. 10a). This suggested that, in addition to Dcr-2, AGO proteins might also participate in hpRNA biogenesis. A role for AGO proteins has also been suggested for the maturation of siRNA duplexes and some pre-miRNAs[2,9–12]. Analysis of mutant animals corroborated this picture of hpRNA biogenesis, because mature hpRNA products were strongly reduced in *Dcr-2*, *loqs* and *AGO2* homozygous mutants (Fig. 2b).

We next analysed the termini of hpRNA-derived small RNAs. β-elimination of RNAs with two free hydroxyl groups at their 3′ termini increases their mobility in denaturing polyacrylamide gel

[1]Sloan-Kettering Institute, Department of Developmental Biology, 521 Rockefeller Research Laboratories, 1275 York Avenue, Box 252, New York, New York 10065, USA. [2]Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, and Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA.

electrophoresis, whereas treatment with calf intestinal phosphatase (CIP) reduces the mobility of 5′ monophosphorylated RNAs[13]. Accordingly, miRNAs run faster after β-elimination and slower after CIP treatment (Fig. 2c and Supplementary Fig. 10b). CIP tests also indicated the presence of 5′ phosphates on hpRNA products (Supplementary Fig. 10b), but all of them were resistant to β-elimination indicating modification of the 3′-terminal ribose (Fig. 2c). *Drosophila* Hen1 methylates Piwi-associated RNAs (piRNAs) and exogenous siRNAs at their 3′ termini[14,15]. We found that *hen1* mutants exhibited lower levels of mature hpRNA products (Fig. 2b), and these were now fully sensitive to β-elimination (Fig. 2c). These data supported the classification of hpRNA products not as miRNAs, but as siRNAs.

We tested the regulatory activity of hpRNA-derived siRNAs using artificial luciferase transcripts linked to target sites that were complementary to various hp-CG4068- and hp-CG18854-derived siRNAs. Their activity was analysed in cells that overexpressed hp-CG4068 or hp-CG18854, with non-cognate pairs controlling for the generic effect of hpRNA overexpression. These tests revealed the specific repression of hp-CG4068B and hp-CG4068C sensors by single- and double-repeat hp-CG4068 expression constructs (Fig. 3a), and of the hp-CG18854B sensor by hp-CG18854 (Fig. 3b). However, a sensor for hp-CG4068D was not affected by ectopic hp-CG4068, consistent with its lower read count compared to hp-CG4068B and hp-CG4068C.

To address the activity of endogenous hpRNAs expressed by S2 cells, we asked whether 2′-O-methyl antisense oligonucleotides (ASOs) could derepress these sensors. Indeed, ASO-hp-CG18854B (but not ASOs to hp-CG18854A or hp-CG4068B) induced approximately twofold derepression of the hp-CG18854B sensor (Fig. 3c). Reciprocally, we observed that ASO-hp-CG4068B (but not other ASOs) resulted in a approximately twofold activation of the hp-CG4068B sensor (Fig. 3d). Thus, both exogenous and endogenous hpRNAs generate inhibitory RNAs.

Some miRNAs are partially loaded into AGO2 (refs 16 and 17), but hpRNA products are the first endogenous *Drosophila* small RNAs known to be preferentially sorted to AGO2 as a class. This provided an opportunity to ask whether endogenously programmed AGO2 functions by means of slicing, translational repression, or both. We prepared hp-CG4068B sensors carrying tandem perfect sites, centrally bulged sites, or bulged plus seed mismatched sites. Both mutant sensors were strongly derepressed, and to roughly the same extent, relative to the perfect sensor (Fig. 3e). In fact, the activity of the

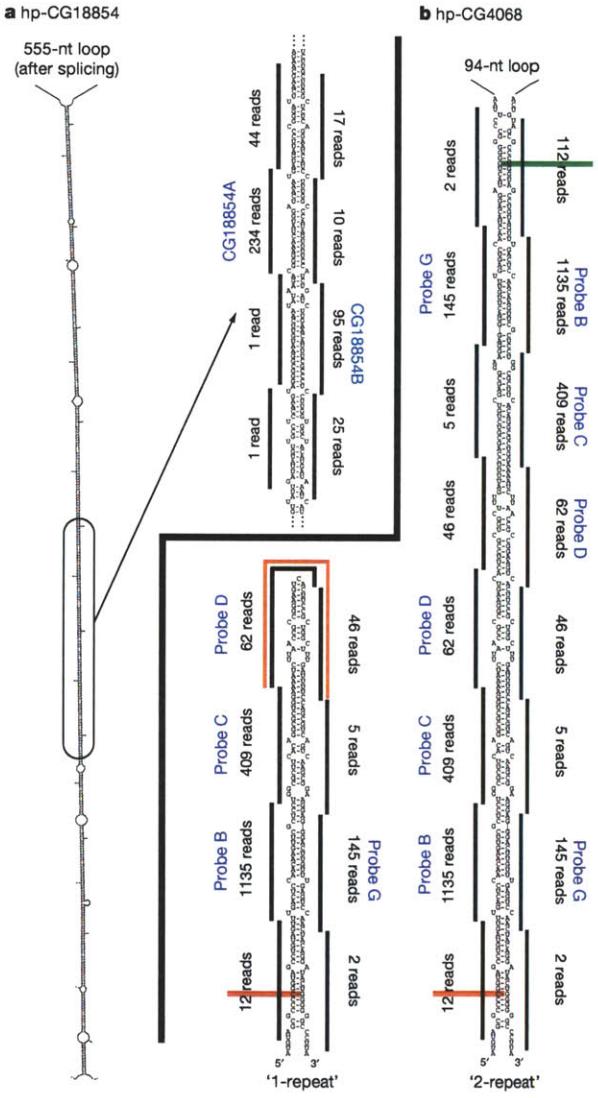


Figure 1 | Examples of *Drosophila* hpRNA transcripts. **a**, hp-CG18854 contains a >400-bp duplex separated by a large loop; the enlarged region highlights the phased nature of small RNA duplexes. Northern probes were designed against the RNAs labelled in blue. **b**, The hp-CG4068 locus consists of 20 tandem repeats that partially overlap the 3′ UTR of CG4068 (Supplementary Fig. 3) and generate phased small RNA duplexes. Each repeat adopts a hairpin structure, but higher-order hairpins are possible because repeats are complementary to each other; '1-repeat' and '2-repeat' isoforms are shown. Distinct small RNAs were cloned from related repeats with minor sequence differences (for example, RNAs highlighted in red and green).

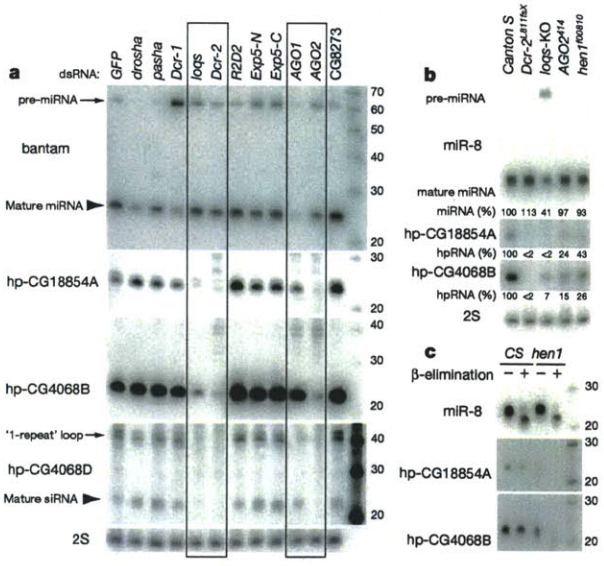


Figure 2 | Distinct biogenesis pathways for miRNAs and hpRNAs. **a**, Unlike miRNAs (for example, Bantam), hpRNA biogenesis in S2 cells is highly dependent on Dcr-2 and AGO2; Loqs and AGO1 suppression affect both miRNA and hpRNA biogenesis. **b**, miRNA and hpRNA biogenesis in pharate adult *Drosophila*. miR-8 was affected only in the *loqs* mutant, whereas hpRNA products were strongly decreased in the *Dcr-2* and *loqs* mutants (<10%), and significantly affected in the *AGO2* and *hen1* mutants. **c**, Modification of hpRNA-derived small RNAs is mediated by Hen1. Note that hpRNA products from *hen1* mutants run as a range of faster-migrating species after β-elimination. CS, Canton S (a strain of fruitfly); KO, knockout deletion strain.

mutant sensors was comparable to the perfect sensor in the presence of cognate ASO (Fig. 3d). These data support the notion that hp-CG4068B exerts its major regulatory effect by slicing, with comparably little contribution from translational repression by AGO2.

With these functional regulatory data in hand, we searched for endogenous targets. We first considered whether hp-CG4068 might regulate the overlapping CG4068 3′ UTR (Supplementary Fig. 3), but gain- and loss-of-function tests were negative (Supplementary Fig. 11). Searches for *trans*-encoded targets were hindered by the fact that hp-CG4068 was only identifiable in the related species *D. melanogaster*, *Drosophila simulans* and *Drosophila sechellia* (Supplementary Fig. 3). However, hp-CG4068B, the most abundant siRNA product of hp-CG4068, contains 20 nucleotides of antisense complementarity (including three G•U pairs) to the coding region of *mus308* (Fig. 3f). This transcript encodes a DNA polymerase/helicase required for DNA repair after exposure to crosslinking agents[18]. Compensatory co-variation between *D. melanogaster* and *D. sechellia* hp-CG4068B and *mus308* target sites were suggestive of functional conservation (Fig. 3f). Consistent with this, we observed that *mus308* levels were increased ~2–3-fold in cells depleted of Dcr-2 or AGO2 (Supplementary Fig. 12), and that a luciferase–*mus308* sensor was specifically repressed by single-repeat and double-repeat hp-CG4068 expression constructs (Fig. 3a).

Because our data indicated that hpRNAs generate functional siRNAs that are primarily dependent on AGO2, we tested whether endogenous hp-CG4068B complexes exhibited Slicer activity. Endogenously programmed complexes indeed cleaved a perfect hp-CG4068B target substrate in a manner that was competed away by ASO-hp-CG4068B but not ASO-CG18854B (Fig. 3g). We also found that a *mus308* target was cleaved by endogenous hp-CG4068B with similar specificity (Fig. 3g). We conclude that *mus308* is an endogenous Slicer target of hp-CG4068B.

We also searched for targets of hp-CG18854. The gene annotated as CG18854 is a possible pseudogene, because its open reading frame is short and poorly conserved[5]. CG18854 exhibits significant homology to the chromodomain gene CG8289 (Supplementary Fig. 13), and some of the abundant hp-CG18854-derived siRNAs exhibited extensive antisense complementarity to CG8289. When tested individually, siRNA-complementary sites from CG8289 did not mediate significant repression (data not shown). We therefore examined whether hp-CG18854 could regulate a translational fusion of CG8289 containing an extended complementary sequence. We transfected S2 cells with either tub-GFP or a tub-CG8289:GFP plasmid along with various hpRNA expression constructs, and observed that hp-CG18854 specifically repressed the accumulation of CG8289:GFP (Supplementary Fig. 13). These data demonstrate that hpRNA products can repress endogenous targets.

In plants, long hairpin RNAs from transgenes and long, extensively paired (and presumably very young) miRNA hairpins are substrates of DICER-LIKE4 (refs 19 and 20), and thus mature through a pathway distinct from that of canonical miRNA hairpins, which are substrates of DICER-LIKE1 (ref. 21). Likewise, we have shown that *Drosophila* hpRNAs enter a pathway distinct from that of miRNAs. Their derivation from unexpectedly long hairpins serves as an important caution for efforts to identify inverted-repeat small RNA genes. For example, some hpRNA-derived clones were recently reported but attributed incorrectly[22], because only short genomic precursors were considered in that study.

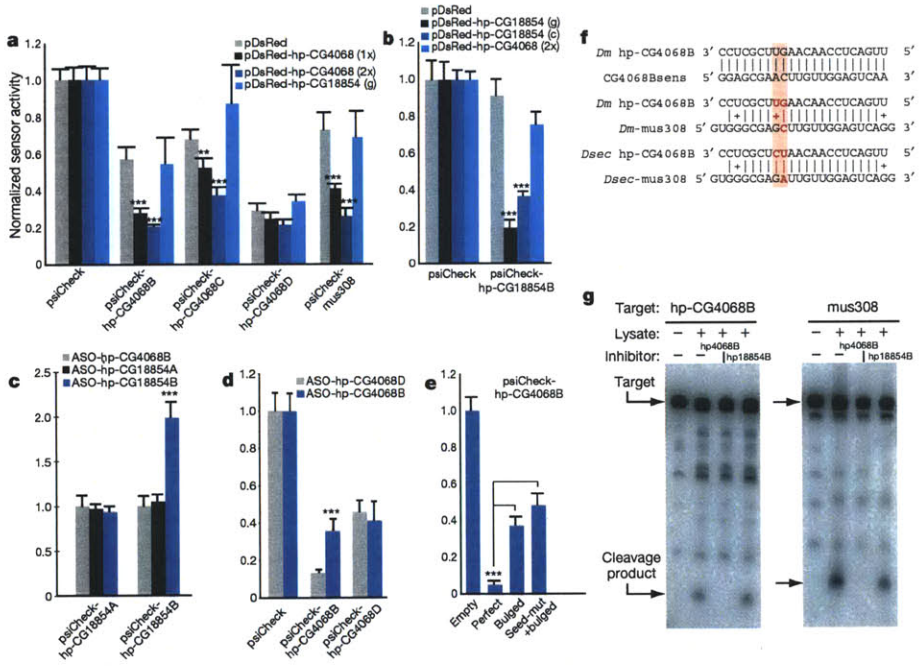


Figure 3 | hpRNAs generate regulatory RNAs that can repress endogenous targets. a, In S2 cells, both single-repeat (1×) and double-repeat (2×) hp-CG4068 constructs specifically repressed the hp-CG4068B, hp-CG4068C and mus308 sensors; the hp-CG4068D sensor was unaffected. **b,** Both genomic (g) and cDNA (c) hp-CG18854 expression constructs specifically repressed the hp-CG18854B sensor. **c,** The 2′-O-methyl ASO against hp-CG18854B specifically derepressed the hp-CG18854B sensor, whereas ASO-CG4068B specifically relieved endogenous repression of the hp-CG4068B sensor (**d**). **e,** Compared to a perfect hp-CG4068B sensor, mutant sensors with a central bulge or central bulge plus seed mutations exhibited the same level of derepression. Error bars depict the standard deviation of eight transfections; statistical comparisons were performed with the unequal variance Students *t*-test; $^{**}P < 6 \times 10^{-5}$, $^{***}P < 1 \times 10^{-8}$. **f,** Compensatory covariation (pink shaded box) between hp-CG4068B and mus308 target sites of *D. melanogaster* (*Dm*) and *D. sechellia* (*Dsec*). Red font, nucleotides that have evolved; sens, sensor. **g,** Endogenous RNA-induced silencing complex from S2R cleaved both perfect hp-CG4068B and *mus308* sensors, and this activity was specifically competed away by cognate ASO.

The *Drosophila* pathway combines canonical RNAi (Dcr-2, Hen1 and AGO2) and miRNA (Loqs) biogenesis factors—a revelation that highlights the incomplete nature of our current understanding of small-RNA-sorting mechanisms. Together with concurrent studies that identify endogenous siRNAs from transposons and *cis*-natural antisense pairs in *Drosophila*[22–24], this work sets the stage for directed studies of the genetic requirements for host-directed RNAi in this organism.

METHODS SUMMARY

We used EINVERTED[4] to identify candidate genomic hairpins contained within 10-kb windows that satisfied a cutoff score ≥80 and had ≥70% pairing within the duplex region. Their expression as small RNAs was analysed using ten 454 libraries[5], a Solexa female head library[23], and a new set of Solexa imaginal disc/brain library (NCBI-GEO GSM275691). We defined candidate hpRNA loci as non-transposon inverted repeats for which the duplex region generated more than three times as many 21–22-nucleotide RNAs than all other-sized RNAs combined (Supplementary Figs 2–9). For functional tests, we followed published protocols for soaking RNAi in S2 cells and northern blotting[25] from knockdown samples or pharate adult flies. For sensor tests, four-copy-site targets (hp-CG4068B, hp-CG4068C and hp-CG4068D sensors) or a two-copy-site target (*mus308* sensor) were prepared by inserting oligonucleotides into a modified version of psiCHECK2 (ref. 25). For hpRNA expression constructs, one or two hp-CG4068 repeats were cloned into the 3′ UTR of UAS-DsRed; CG18854 fragments were amplified from genomic DNA (CG18854 genomic) or a LD34273 clone (CG18854 cDNA) and were cloned similarly. A CG8289:GFP translational sensor consisted of a CG8289 fragment from genomic DNA and cloned into the KpnI site of tub-GFP[26]. RNA 3′ termini were analysed using periodate treatment in borax/boric-acid buffer followed by NaOH treatment (β-elimination) as described[15]. RNA 5′ termini were analysed using CIP as described[13]. For cleavage assays, we prepared labelled gel-purified templates using α-^{32}P-GTP and capping enzyme (Ambion). Cleavage reactions were performed as described[2] using S2-R+ cell extract. For detailed bioinformatic and molecular methods, see Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 January; accepted 22 April 2008.
Published online 7 May 2008.

1. Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
2. Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).
3. Kennerdell, J. R. & Carthew, R. W. Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nature Biotechnol.* **18**, 896–898 (2000).
4. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
5. Wilson, R. J., Goodman, J. L. & Strelets, V. B. FlyBase: integration and improvements to query tools. *Nucleic Acids Res.* **36**, D588–D593 (2008).
6. Tupy, J. L. *et al.* Identification of putative noncoding polyadenylated transcripts in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **102**, 5495–5500 (2005).
7. Jones-Rhoades, M. W., Bartel, D. P. & Bartel, B. MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**, 19–53 (2006).
8. Liu, X. *et al.* Dicer-1, but not Loquacious, is critical for assembly of miRNA-induced silencing complexes. *RNA* **13**, 2324–2329 (2007).
9. Leuschner, P. J., Ameres, S. L., Kueng, S. & Martinez, J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep.* **7**, 314–320 (2006).
10. Matranga, C., Tomari, Y., Shin, C., Bartel, D. P. & Zamore, P. D. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607–620 (2005).
11. Diederichs, S. & Haber, D. A. Dual role for Argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell* **131**, 1097–1108 (2007).
12. Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
13. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).
14. Saito, K. *et al.* Pimet, the *Drosophila* homolog of HEN1, mediates 2′-O-methylation of Piwi- interacting RNAs at their 3′ ends. *Genes Dev.* **21**, 1603–1608 (2007).
15. Horwich, M. D. *et al.* The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* **17**, 1265–1272 (2007).
16. Seitz, H., Ghildiyal, M. & Zamore, P. D. Argonaute loading improves the 5′ precision of both microRNAs and their miRNA strands in flies. *Curr. Biol.* **18**, 147–151 (2008).
17. Forstemann, K., Horwich, M. D., Wee, L., Tomari, Y. & Zamore, P. D. *Drosophila* microRNAs are sorted into functionally distinct Argonaute complexes after production by Dicer-1. *Cell* **130**, 287–297 (2007).
18. Harris, P. V. *et al.* Molecular cloning of *Drosophila mus308*, a gene involved in DNA cross-link repair with homology to prokaryotic DNA polymerase I genes. *Mol. Cell. Biol.* **16**, 5764–5771 (1996).
19. Dunoyer, P., Himber, C. & Voinnet, O. DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nature Genet.* **37**, 1356–1360 (2005).
20. Rajagopalan, R., Vaucheret, H., Trejo, J. & Bartel, D. P. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev.* **20**, 3407–3425 (2006).
21. Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B. & Bartel, D. P. MicroRNAs in plants. *Genes Dev.* **16**, 1616–1626 (2002).
22. Ghildiyal, M. *et al.* Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* doi:10.1126/science.1157396; published online 10 April 2008.
23. Czech, B. *et al.* An endogenous small interfering RNA pathway in *Drosophila*. *Nature* doi:10.1038/nature07007 (this issue).
24. Kawamura, Y. *et al.* *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* doi:10.1038/nature06938 (this issue).
25. Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M. & Lai, E. C. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **130**, 89–100 (2007).
26. Stark, A., Brennecke, J., Russell, R. B. & Cohen, S. M. Identification of *Drosophila* microRNA targets. *PLoS Biol.* **1**, E60 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to R. Carthew, Q. Liu, H. Siomi and S. Cohen for plasmids and *Drosophila* strains. K.O. was supported by the Charles H. Revson Foundation. H.G. was supported by A*STAR, Singapore. D.P.B. is an HHMI investigator, and work in his laboratory was supported by a grant from the NIH (GM067031). E.C.L. was supported by grants from the Leukemia and Lymphoma Foundation, the Burroughs Wellcome Foundation, the V Foundation for Cancer Research, the Sidney Kimmel Foundation for Cancer Research, and the NIH (GM083300).

Author Contributions J.G.R. identified hp-CG4068 and hpRNA1. W.-J.C. performed the EINVERTED analysis and identified the additional hpRNA loci and their targets. H.G. performed initial hpRNA northern analysis; all other experiments were designed and carried out by K.O. All authors contributed to the preparation of the manuscript.

Author Information The imaginal disc/brain sample described in paper has been deposited in the NCBI GEO under accession number GSM275691. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.C.L. (laie@mskcc.org).

METHODS

Bioinformatics. We used EINVERTED[4] to identify candidate genomic hairpins contained within 10-kb windows that satisfied a cutoff score ≥80 and had ≥70% pairing within the duplex region. These criteria eliminated all but one of the annotated *Drosophila* miRNAs (*mir-997*)[27]. We kept those candidates with small RNA reads in the following data sets: 10 libraries analysed using 454 pyrosequencing[5], a female head library analysed using Solexa[23], and a new set of imaginal disc/brain small RNA sequences analysed using Solexa. We removed those loci in which the predicted duplex overlapped an annotated transposable element[28], and calculated the size distribution of reads from each of the remaining loci. We considered those loci for which the duplex region generated more than three times as many 21–22-nucleotide RNAs than all other-sized RNAs combined as hpRNA candidates (Supplementary Figs 2–9).

RNA interference. Segments of Pasha, Drosha and CG8273 were amplified using the indicated primers and were cloned into XhoI-XbaI sites of Litmus28i vector (NEB); other plasmids were described previously[25,29]. These templates were used to generate dsRNA, and soaking RNAi was performed as described[25]. S2-R+ cells were resuspended in serum-free medium at 3×10^6 cells ml^{-1} density and dsRNA was added to a concentration of 15 μg ml^{-1}. After 30 min incubation, an equal volume of Schneider's medium supplemented with 20% FBS was added. dsRNA treatment was repeated 4 days after the first treatment, and RNA samples were collected 4 days after the second soaking. XhoI-CG8273-279+, AGAGCTCGAGTCAGACAAATCCTCCGGTTC; XbaI-CG8273-682−, AGAGGTCTAGATTCGCCATCTGACTTGGTC; XhoI-Pasha-1240+, AGAGCTCGAGGGAGGTGGAGCAACAAAAGA; XbaI-Pasha-1725−, AGAGGTCTAGATACTCGTGCAGGATGCAGAC; XhoI-Drosha3522+, AGAGCTCGAGGCCGGACATTCCCTACTACA; XbaI-Drosha3943−, AGAGGTCTAGAGGGCATTGTTGGACTCTTG.

Northern blotting. Northern blotting was performed as described[25] using total RNA isolated from S2 cells or pharate adult *Drosophila*. All of the mutant strains used were described previously: *Dcr-2*^{L811fsX} (ref. 1), *loqs*-KO (ref. 30), *AGO2*^{414} (ref. 2) and *hen1*^{f00810} (refs 14 and 15). The sequences of the probes used in this study are listed below. DNA and locked nucleic acid (LNA) probes were obtained from IDT and Exiqon, respectively. *bantam* probe (DNA), AATCAGCTTTCAAAATGATCTCA; 2S rRNA probe (DNA), TACAACCCTCAACCATATGTAGTCCAAGCA; hp-CG4068-D (LNA), GTGACTTCCGGCGGTTAAGATTT; hp-CG4068-B (LNA), GGAGCGAACTTGTTGGAGTCAA; hp-CG4068-G (LNA), AGTTGGACTCAAACAAGTCCCT; hp-CG18854-A (LNA), TCATTTGATCCATAGTTTCCCGT; hp-CG18854-B (LNA), GGAGGGCGAAATGTTCAAGATCA; miR-8 (LNA), A.

Analysis of RNA chemical structure. RNA 3′ termini were analysed as described[15]. 10 μl of 61.5 mM $NaIO_4$ in borax/boric-acid buffer (60 mM borax and 60 mM boric acid, pH 8.6) was added to 10 μg total RNA in 14.6 μl water, and the samples were incubated for 30 min at room temperature (22 °C). 2.5 μl of 500 mM NaOH was added to each sample and incubation was continued for 90 min at 45 °C. The reactions were stopped by addition of 200 μl of 300 mM NaCl, 10 μg glycogen and 600 μl absolute ethanol. RNA was collected by centrifugation after 30 min incubation on ice.

RNA 5′ termini were analysed as described[13]. RNA samples were incubated with 2 units CIP (New England Biolabs) in 1× buffer 3 (NEB, 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol) for 2 h at 37 °C. RNA was purified by phenol/chloroform extraction followed by ethanol precipitation.

Sensor assays. Four-copy-site targets (hp-CG4068B, hp-CG4068C and hp-CG4068D sensors) or a two-copy-site target (*mus308* sensor) were prepared using the oligonucleotides listed below. Target sequences were inserted into SalI-XhoI (four-copy sensors) or NotI-XhoI (two-copy sensor) cloning sites of a modified version of psiCHECK2 (ref. 25). CG4068 3′ UTR sensor was constructed by inserting a CG4068 3′ UTR fragment amplified with primers CG4068A and B (containing one repeat of the hpRNA repeat) into NotI-XhoI sites of the modified psiCHECK2. CG18854 fragments were amplified from genomic DNA (CG18854 genomic) or LD34273 clone (CG18854 cDNA) and cloned into NotI-XhoI sites of UAS-DsRed. The CG8289 fragment was amplified from genomic DNA and cloned into KpnI site of tub-GFP plasmid[26].

Luciferase sensor assays were performed as described previously[25]. We performed quadruplicate transfections of 25 ng target, 12.5 ng ub-Gal4 and 25 ng UAS-DsRed-hpRNA plasmids into 1×10^5 S2 cells in 96-well format. For 2′-O-methyl antisense-mediated de-silencing assays (inhibitor sequences listed below), we introduced 25 ng target plasmid and 10 pmol of 2′-O-methyl oligonucleotides for each well. Three days later, we lysed the cells and subjected them to the dual luciferase assay (Promega) and analysed these on a Veritas plate luminometer (Turner Biosystems). KpnI-CG8289 targetF, ggggtaccgccgccaccatgGTTGCTGAAAAGGATTCG; KpnI-CG8289, ggggtaccTTCGAGGAGCGTTCAATACGAT targetR; NotI-LD34273-1+, AGAGgcggccgcAGTGCTGAGCATACCTAAGC; XhoI-LD34273-2390−, AGAGctcgagGTTCCCACATCGACTGGAAT; CG4068_A, aggggcggccgcACAAGCCAAAATCGTAtagg; CG4068_B, agggctcgagTTTTGCGTGGACTCATTCCC; hp-CG4068B_targ_A, tcgacaaaaGGAGCGAACTTGTTGGAGTCAAaagaac; hp-CG4068B_targ_B, tcgagttcttTTGACTCCAACAAGTTCGCTCCttttg; hp-CG4068C_targ_A, tcgacaaaaTTCCAGCGCCTGTGAAGCGCCAgagaac; hp-CG4068C_targ_B, tcgagttctcTGGCGCTTCACAGGCGCTGGAAttttg; hp-CG4068D_targ_A, tcgacaaaaGTGACTTCCGGCGGTTAAGATTTagaac; hp-CG4068D_targ_B, tcgagttctAAATCTTAACCGCCGGAAGTCACttttg; hp-CG4068A_si2x_A, GGCCGCGGAGCGAACTTGTTGGAGTCAAaaatcacGGAGCGAACTTGTTGGAGTCAAaC; hp-CG4068A_si2x_B, TCGAGtTTGACTCCAACAAGTTCGCTCCgtgatttTTGACTCCAACAAGTTCGCT GC; hp-CG4068A_mi2x_A, GGCCGCGGAGCGAACTACATCCACTGAAaaatcacGGAGCGAACTACATCCACTGAAaC; hp-CG4068A_mi2x_B, TCGAGtTTCAGTGGATGTAGTTCGCTCCgtgatttTTCAGTGGATGTAGTTCGCT GC; hp-CG4068A_mimut2x_A, GGCCGCGGAGCGAACTACATGGAGTCAAaaatcacGGAGCGAACTACATGGAGTCAAaC; hp-CG4068A_mimut2x_B, TCGAGtTTGACTCCATGTAGTTCGCTCCgtgatttTTGACTCCATGTAGTTCGCT GC; mus308-target1not, ggccatggGCGAGCTTGTTGGAGTCAgggtgatttggGCGAGCTTGTTGGAGTCAggc; mus308-target2, tcgagcctgactccaacaagctcgcccaaatcacccTGACTCCAACAAGCTCGCcat; ASO-hp-CG4068-B, 5′AACAUggagcgaacttgttggagtcaaAUACA 3′; ASO-hp-CG4068-C, 5′AACAUttccagcgcctgtgaagcgccagUCACU 3′; ASO-hp-CG4068-D, 5′AACAUGUGACUUCCGGCGGUUAAGAUUUAUACA 3′; ASO-hp-CG18854-A, 5′AACAUtgGCCAAGGTACGTGGTCGACCGAAUACU 3′; ASO-hp-CG18854-B, 5′AACAUGGAGGGCGAAATGTTCAAGATCAUCACU 3′.

For the green fluorescent protein (GFP) sensor assay, 250 ng target, 125 ng ub-Gal4 and 250 ng UAS-DsRed-hpRNA plasmids were transfected to 2×10^6 cells in 6-well format. Three days later, transfected cells were harvested and lysed with 2× SDS–PAGE sample buffer. Western blotting was performed using rabbit anti-GFP (Molecular Probes) or mouse anti-α-tubulin (DM1A, Sigma).

***In vitro* cleavage assay.** Templates for *in vitro* cleavage targets were prepared by treating annealed oligonucleotides with Taq polymerase. The oligonucleotides A and B or A and C (below) were used for the template preparation for hp-CG4068-B target or mus308 target, respectively. Target RNAs were *in vitro* transcribed using Megascript T7 kit (Ambion) and purified by acrylamide gel electrophoresis. Purified RNA was labelled by α-^{32}P-GTP using capping enzyme (Ambion) according to the manufacturer's instructions. The cleavage reaction was performed as described[2] using S2-R+ cell extract. S2-R+ cells were resuspended in hypotonic buffer (30 mM HEPES-KOH, pH 7.4, 2 mM magnesium acetate, 5 mM DTT, 1× Complete mini EDTA free (Roche)) and lysed by five passages through a 25-gauge needle. The lysate was cleared by a centrifugation for 25 min at 14,600*g* at 4 °C, and was flash-frozen in 10-μl aliquots. Approximately 2,000 counts per min cap-labelled RNA was incubated in a reaction mixture (50% S2 lysate, 0.5 mM ATP, 5 mM DTT, 100 mM KOAc, 0.1 U μl^{-1} RNaseOut (Invitrogen)) for 1 h at room temperature. 2′-O-methyl-ASO inhibitors (Integrated DNA Technologies) were added at 100 nM concentration and preincubated with the reaction mixture at room temperature for 20 min before the addition of the cap-labelled target RNA. Reactions were stopped by the addition of stop buffer (50 mM sodium chloride, 50 mM EDTA, 1% SDS, and 100 μg ml^{-1} proteinase K). RNA was recovered by phenol/chloroform extraction and ethanol precipitation. A, LucLet7_3′region_AS, attaattcctatGAGGTAGTAGGTTGTATAGTTCGAAGTATTCCGCGTACGTT; B, T7_hp-CG40682B_Luc_sense, TAATACGACTCACTATAGGGAGCGAACTTGTTGGAGTCAAattcAACGTACGCGGAATAC; C, T7_Mus308tar_Luc_sense, TAATACGACTCACTATAGggtggGCGAGCTTGTTGGAGTCAgggAACGTACGCGGAATAC.

27. Ruby, J. G. *et al.* Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* **17**, 1850–1864 (2007).
28. Karolchik, D. *et al.* The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.* **36**, D773–D779 (2008).
29. Forstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3**, e236 (2005).
30. Park, J. K., Liu, X., Strauss, T. J., McKearin, D. M. & Liu, Q. The miRNA pathway intrinsically controls self-renewal of *Drosophila* germline stem cells. *Curr. Biol.* **17**, 533–538 (2007).

Curriculum Vitae

Huili Guo

Education

2006 – 2011	Ph.D. Biology Massachusetts Institute of Technology
2002 – 2005	B.A. First Class Honours (Natural Sciences – major in Biochemistry) Sidney Sussex College, University of Cambridge, UK

Research Experience

2007 – 2011	Graduate studies, Laboratory of Prof. David P. Bartel, MIT A study of mammalian microRNA-mediated repression of gene expression by ribosome profiling
2005 –2006	Research Assistant, Laboratory of A/Prof. Graeme R. Guy Institute of Molecular and Cell Biology (IMCB), Singapore The role of Spred in small G-protein signaling
Spring 2005	Undergraduate thesis project, Laboratory of Dr. Richard W. Farndale, Department of Biochemistry, Cambridge, UK Activation of platelet receptors by collagen mimetics
Summer 2004	Research project, Laboratory of Prof. James C. Smith Gurdon Institute of Cancer and Developmental Biology, Wellcome Trust/Cancer Research UK (grant from Nuffield Foundation) Functional analysis of T-box transcription factors in *Xenopus tropicalis*
Summer 2003	Research project, Laboratory of A/Prof. Graeme R. Guy, IMCB, Singapore Characterization of a splice variant of the mErk5 protein

Teaching Experience

Spring, 2010	Teaching Assistant, Molecular Biology (7.28/7.58), MIT
Fall, 2007	Teaching Assistant, Experimental Biology and Communication (7.02), MIT

Awards

2010	RNA Society Poster Award, Gordon Research Conference, on the Biology of Post-Transcriptional Gene Regulation, Newport, RI
2003 – 2005	Elected to Samuel Taylor Scholarship for excellence in Natural Sciences Tripos, Sidney Sussex College, Cambridge, UK
2004	Nuffield Foundation Undergraduate Research Bursary, to perform an eight-week research project (see above)
2004	Jacks Prize, for academic distinction in Natural Sciences Tripos
2003	Graham, Day, Bellerby Prize, for academic distinction in Natural Sciences Tripos
2002	National Science Scholarship, Agency for Science, Technology and Research (A*STAR), Singapore

Publications

Guo, H., Ingolia, N. T., Weissman, J. S., Bartel, D. P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466: 835-840

Okamura, K., Chung, W.-J., Ruby, G. R., **Guo, H.**, Bartel, D. P., Lai, E. C. (2008) The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature 453: 803-806

McCaw, B.J., Chow, S.Y., Wong, E.S., Tan, K.L., **Guo, H.**, Guy, G.R. (2005) Identification and characterization of mErk5-T, a novel Erk5/Bmk1 splice variant. Gene 345: 183-190